

Программные системы и вычислительные методы

Правильная ссылка на статью:

Скляр А.Я. Модели и алгоритмы нелинейного регрессионного анализа временных рядов // Программные системы и вычислительные методы. 2025. № 3. DOI: 10.7256/2454-0714.2025.3.75317 EDN: OPDYKY URL: https://nbpublish.com/library_read_article.php?id=75317

Модели и алгоритмы нелинейного регрессионного анализа временных рядов

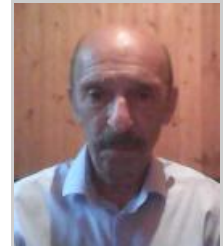
Скляр Александр Яковлевич

кандидат технических наук

доцент, кафедра прикладная математика; Российский технологический университет (МИРЭА)

119602, Россия, г. Москва, Вернадского, 78

✉ askliar@mail.ru



[Статья из рубрики "Системный анализ, поиск, анализ и фильтрация информации"](#)

DOI:

10.7256/2454-0714.2025.3.75317

EDN:

OPDYKY

Дата направления статьи в редакцию:

27-07-2025

Аннотация: Анализ данных, описывающих те или иные объекты и процессы предназначен, прежде всего, для нахождения зависимостей внутри них и выявления динамики их развития. Целями анализа и прогнозирования является подготовка материалов для принятия обоснованных решений. В данной работе рассматриваются этапы, методы и алгоритмы проведения анализа, направленные на получение, в первую очередь, функциональных зависимостей, пригодных не только для описания, но и прогнозирования поведения изучаемых объектов и процессов. Сам анализ рассматривается как многоэтапный процесс, включающий подготовку данных, выявление и удаление по возможности шума в данных, нахождение долгосрочного тренда, выявление колебательной составляющей, периодичности колебаний, оценку динамики амплитуд колебаний, оценка точности возможной аппроксимации процесса и возможности его прогнозирования с учетом уровня зашумленности данных. Предложен ряд процедур, обеспечивающих обоснованную проверку гипотез о ходе процессов и получение аналитических, в том числе и дифференциальных зависимостей на основе методов оптимизации подбора параметров нелинейных зависимостей. Рассмотренные

методы позволяют проводить достаточно объективный анализ данных и создают условия для построения обоснованного прогноза. Проведен численный анализ по данным многолетней статистики динамики производства. Научная новизна исследования заключается в разработке методики декомпозиции процесса на трендовые и колебательные компоненты. В отличие от большинства существующих исследований анализа динамики процессов большое внимание уделяется учету и оценке уровня шума с определением пределов точности получаемых результатов и, тем более, прогнозов, что позволяет избегать необоснованных выводов и решений и построения «слишком» точных результатов на основе недостаточно точных исходных данных, исходя из требований гладкости функции при имеющемся уровне шума. Использование функций ограниченного роста и выявление точек смены тренда позволяет проводить корректное качественное долгосрочное прогнозирование без необоснованного предсказания катастрофического хода изучаемого процесса. Полученные результаты позволяют получить аналитическое выражение изучаемых, прежде всего экономических процессов, что позволяет не только аппроксимировать поведение процесса, но и выявлять его физическую сущность и, соответственно использовать полученные решения на целый класс процессов аналогичной природы.

Ключевые слова:

анализ данных, временной ряд, удаление шума, долгосрочный тренд, оценка точности, аппроксимация данных, нелинейная регрессия, проверка гипотез, ограниченный рост, анализ колебаний

Введение

Анализ статистических данных предназначен, прежде всего для выявления внутренних зависимостей. Применительно к данным, представленным во временном ряде – это выявление динамики развития характеристики процесса, описываемого этими данными. Большинство процессов, прежде всего экономических, являются непрерывными. Их характеристики могут меняться в достаточно широких, но всегда ограниченных диапазонах. Сами данные во временных рядах, описывающих процесс дискретны и подвержены влиянию случайных факторов начиная с ошибок измерения и различных внешних неконтролируемых возмущений. В этих условиях задача анализа распадается на ряд поэтапно решаемых подзадач.

Выделим основные этапы:

удаление шума (случайных возмущений);

нахождение долговременного тренда;

выявление колебательной составляющей;

оценка динамики амплитуд колебаний;

оценка точности возможной аппроксимации процесса и возможности его прогнозирования с учетом уровня зашумленности данных; следует всегда помнить, что точность прогноза ни при каких обстоятельствах не может быть выше уровня шума.

В случаях, когда процессы описываются набором дискретных состояний, обработка должна строиться на выделении точек перехода с использованием иного инструментария.

К основным методам, широко используемым в настоящее время, относятся методы, основанные на использовании деревьев решений [1, 2], кластерного анализа [3], нейронных сетей [4]. Основы этих методов описаны и в учебной литературе, например, в учебнике М.Ю. Мошкова (Мошков М. Ю. Деревья решений. Теория и приложения. Нижний Новгород. Изд. Нижегородского ун. 1994. 176 с.).

Оценка и удаление шума в данных

Анализ и обработка зашумленных данных вызывает значительные трудности. Возникает задача устранения, по возможности, такого шума. Для устранения такого шума используются различные методы сглаживания, такие, как методы скользящей средней, экспоненциального сглаживания и др. (Эконометрика. Учебник / Под ред. Елисеевой И. И. — 2-е изд. — М.: Финансы и статистика, 2006. — 576 с. — ISBN 5-279-02786-3), а также методы полиномиального сглаживания, прежде всего Савицкого – Галея [5].

Обозначим значения аргумента (временные отрезки) через x_i , соответствующие им наблюдаемые значения через y_i , предполагаемую функцию «истинной» зависимости – $f(x)$. Далее будем рассматривать случай, когда исходные данные задаются с постоянным шагом h и $x_{i+1}=x_i+h$.

Тогда наблюдаемые данные можно представить в виде $y_i=f(x_i)+s_i$, где s_i – шум.

Функцию $f(x)$ будем считать k раз дифференцируемой, тогда ее производная может быть вычислена как

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - r$$

Здесь r – систематическая ошибка

$$r = \frac{h^2}{3!} f'''(x) + \dots + \frac{h^{2k}}{(2k+1)!} f^{(2k+1)}(x + \theta h)$$

Взяв значения функции в нескольких последовательных точках, можно показать, что всегда существует набор величин λ_i таких, что

$$\sum_{i=-t}^{k+t-1} \lambda_i f(x_{m+i}) = O(f^{(k-1)}(x_m) \Delta x^{k-1}) \quad (1)$$

В частности, при равномерном шаге при $k=3, 4, 5$ получим наборы λ_i (1, -2, 1), (-1, 3, -3, 1), (1, -4, 6, -4, 1) соответственно, что соответствует численным выражениям для 2, 3 и 4 производных [6, 7]. Отклонение от 0 таких сумм для любых слабо осциллирующих функций пренебрежимо мало.

Рассмотрим

$$\sum_{i=-t}^{k+t-1} \lambda_i y_{m+i} = \sum_{i=-t}^{k+t-1} \lambda_i f(x_{m+i}) + \sum_{i=-t}^{k+t-1} \lambda_i s_{m+i} \approx \sum_{i=-t}^{k+t-1} \lambda_i s_{m+i} \quad (2)$$

Величины s_i – независимые случайные числа с нулевым математическим ожиданием.

Среднеквадратичное значение σ для s можно получить

$$\left(\sum_{i=-t}^{k+t-1} \lambda_i y_{m+i} \right)^2 \approx \left(\sum_{i=-t}^{k+t-1} \lambda_i s_{m+i} \right)^2 = \sigma^2 \sum_{i=-t}^{k+t-1} \lambda_i^2 \quad (3)$$

При 5 точечной схеме это дает с равномерным шагом это дает значение $70\sigma^2$. Просуммировав эти величины по n точкам ряда, получим значение D , где $D=70n\sigma^2$, откуда $\sigma^2=D/70n$. Таким образом получаем объективную оценку уровня шума в данных.

Определить значения s_i , а значит и f_i можно решением

$$F(s) = \sum_{m=3}^{n-2} \left(\sum_{k=-2}^2 \lambda_{k,m} (y_{k+m} - s_{k+m}) \right)^2 \rightarrow \min; \sum_{k=1}^n s_i^2 = n\sigma^2 \quad (4)$$

Полученное решение дает сглаженный ряд, представляющий максимально близкую к исходному гладкую функцию, что позволяет использовать не только входные данные, но и их производные [6, 7].

Нахождение долговременного тренда

Говоря о тренде $u(t)$, обычно понимают некоторое усреднение данных $f(t)$ в той или иной мере приближающееся к конструкции $u(t) = \frac{1}{T} \int_{t-T/2}^{t+T/2} f(x) dx$. Здесь T задает промежуток усреднения, определяя характер тренда. На практике обычно применяют методики скользящего среднего или многократного скользящего выравнивания. Проблемой подобных методов помимо их зависимости от способа выравнивания является отсутствие сколько-нибудь явной аналитической зависимости.

Альтернативой такому подходу является проверка соответствия исходных данных некоторой теоретической кривой $u(x, p_1, \dots, p_k)$ с набором параметров. Задача состоит в подборе этих параметров на основе, например, метода наименьших квадратов.

$$\sum (y(x_i) - u(x_i, p_1, \dots, p_k))^2 \rightarrow \min$$

Эта задача сводится к решению системы уравнений

$$\sum (y(x_i) - u(x_i, p_1, \dots, p_k)) \frac{\partial u}{\partial p_j} = 0$$

В общем случае это система из k нелинейных уравнений, решение которых достаточно затруднительно. Система становится линейной только, когда функция u линейна по всем своим параметрам, например, когда u полином от x . Такая конструкция может с какой угодно точностью аппроксимировать исходные данные, но «дальний» прогноз на ее основе теряет смысл, поскольку он в зависимости от знака старшего коэффициента стремится либо к $+\infty$, либо к $-\infty$.

В ряде случаев преобразование системы координат (анаморфоза) позволяет выстроить исследуемые данные в линейную зависимость. Например, переход в логарифмические или полулогарифмические координаты линеаризует экспоненциальную, степенную и логарифмическую зависимости [8]. Последнее, к сожалению, возможно лишь когда целевая переменная не меняет свой знак.

Более эффективные результаты можно получить, если данные исходного ряда дифференцируемы, что может быть достигнуто, например, описанной выше процедурой удаления шума.

Для природных и, прежде всего, экономических процессов, функционирующих в относительно постоянных условиях, характерно асимптотическое стремление к некоторому уровню. Такие процессы естественным образом описываются моделями ограниченного роста. Рассмотрим такие процессы и методы определения их параметров.

Дифференциальные методы.

Экспоненциальное приближение к постоянной $y = a + be^{-kt}$.

$$y' = -kbe^{-kt} = -k(y - a) = \left| k^* = \frac{1}{k} \right|; k^* y' - a = -y \quad (5)$$

Здесь параметры a и k^* входят линейно и, подставляя исходные значения ряда в y и y' методом наименьших квадратов находим требуемые параметры, причем знак и наличие колебаний в исходных данных не являются препятствием. Зная параметры a и k из соотношения $y = a + be^{-kt}$ находим параметр b .

Гиперболическое приближение к постоянной $y = a + b/(x+c)$.

$$y' = -\frac{b}{(x+c)^2}; y'' = 2\frac{b}{(x+c)^3} = \frac{-2y'}{x+c}; y''x + cy'' = -2y' \quad (6)$$

Отсюда последовательно находим параметр c и далее a и b .

Либо без использования высших производных

$$y' = -\frac{b}{(x+c)^2} = -\frac{1}{x+c} \frac{b}{x+c} = \frac{-(y-a)}{x+c}; cy' - a + (xy' + y) = 0 \quad (7)$$

Модель Гомперца $y'/y = ae^{-kx}$; $y = y_{\infty} e^{ay - kx}$

$$u = \frac{y'}{y}; u = ae^{-kx}; u' = -kae^{-kx} = -ku \quad (8)$$

Отсюда последовательно находим параметр k и далее a , и y_{∞} .

Сигмоидальные модели $y = a + \frac{b}{e^{-k(x-c)} + 1}$; $y = a + \frac{b(x-c)}{k|x-c|+1}$; $y = a + \frac{b(x-c)}{\sqrt{k^2(x-c)^2+1}}$; $y = a + \text{barctg}(k(x-c))$ и т.п. с тем свойством, что их графики обладают центральной симметрией относительно точки $(c, y(c))$.

Точку c можно найти минимизацией $f(c)$

$$f(c) = \frac{1}{h} \sum_{i=1}^h ((y_{c+i} - y_c^*) - (y_c^* - y_{c-i}))^2 = \frac{1}{h} \sum_{i=1}^h (y_{c+i} + y_{c-i} - 2y_c^*)^2 \quad (9)$$

В качестве y_c^* можно принять $y_c^* = y_c$ или предпочтительней $y_c^* = \frac{1}{2h+1} \sum_{i=-h}^h y_{c+i}$

После этого переход к $z = y - y_c^*$ сигмоиду с нулем в точке c .

Так для классической экспоненциальной сигмоиды получим $z = bth(k(x-c))$.

$$z' = bk \frac{ch^2(k(x-c)) - sh^2(k(x-c))}{ch^2(k(x-c))} = bk(1 - th^2(k(x-c))) = bk(1 - \left(\frac{z}{b}\right)^2)$$

$$\frac{1}{bk} z' = 1 - \frac{1}{b^2} z^2 \quad (10)$$

И введя $p=1/bk$, $q=1/b^2$ получим $pz' + qz^2 - 1 = 0$. Откуда методом наименьших квадратов находятся параметры p и q и, следовательно a и b .

Интегральные методы.

В ряде случаев дифференциальные методы плохо применимы из-за высокого уровня погрешностей при вычислении производных, особенно высших порядков. В то же время аналогичный подход вполне реализуем при замене производных соответствующими интегралами.

Рассмотрим экспоненциальное приближение к постоянной $y=a+be^{-kt}$.

Введем

$$Y = \int_0^x y dx = \int_0^x (a + be^{-kx}) dx = ax + \frac{b}{k} - \frac{1}{k} be^{-kx} = ax + \frac{b}{k} - \frac{1}{k}(y - a) \quad (11)$$

$$Y = ax + \frac{b+a}{k} - \frac{1}{k}y = ax + c - k^*y; k = \frac{1}{k^*}; b = ck - a \quad (12)$$

Получение Y численным интегрированием исходных данных не вызывает никаких проблем и легко реализуется стандартными методами, например средствами Excel и следом выполнив нахождение параметров a , k и b методом наименьших квадратов получим требуемую линию тренда.

Аналогично можно построить и алгоритмы для других гипотез для нахождения линий тренда. Существенным здесь является тот факт, что данная методика применима для любых данных вне зависимости от наличия в них колебаний.

Характеристические показатели

Параметры моделей ограниченного роста могут быть оценены на основе набора характеристических интегральных показателей выдвигаемой гипотезы.

Экспоненциальное приближение к постоянной

Для унификации расчетов линейно преобразуем интервал изменения аргумента к интервалу [\[0, 1\]](#).

Простая экспонента

$$y = be^{-kt} \quad (t \in [0, 1])$$

$$S_0 = \int_0^1 be^{-kt} dt = -\frac{b}{k} e^{-kt} \Big|_0^1 = \frac{b}{k} (1 - e^{-k})$$

$$S_1 = \int_0^1 bte^{-kt} dt = -b \left(\frac{x}{k} + \frac{1}{k^2} \right) e^{-kt} \Big|_0^1 = \frac{b}{k} \left(\frac{1}{k} - \left(1 + \frac{1}{k} \right) e^{-k} \right)$$

$$L = \frac{S_1}{S_0} = \frac{1 - e^{-k}}{\frac{1}{k} - \left(1 + \frac{1}{k} \right) e^{-k}} \quad (13)$$

Значения $L(k)$ табулируются. После этого нахождение $k(L)$ не вызывает проблем. Далее находим $b = S_0 k(L) / (1 - e^{-k})$

Смещенная экспонента

$$y = a + be^{-kt} \quad (t \in [0, 1])$$

$$S_0 = \int_0^1 (a + be^{-kt}) dt = at - \frac{b}{k} e^{-kt} \Big|_0^1 = a + \frac{b}{k} (1 - e^{-k})$$

$$S_1 = \int_0^1 (a + be^{-kt}) t dt = \frac{at^2}{2} - b \left(\frac{t}{k} + \frac{1}{k^2} \right) e^{-kt} \Big|_0^1 = \frac{a}{2} + \frac{b}{k} \left(\frac{1}{k} - \left(1 + \frac{1}{k} \right) e^{-k} \right)$$

$$S_2 = \int_0^1 (a + be^{-kt}) t^2 dt = \frac{at^3}{3} - b \left(\frac{t^2}{k} + \frac{2t}{k^2} + \frac{2}{k^3} \right) e^{-kt} \Big|_0^1 = \frac{a}{3} + \frac{b}{k} \left(\frac{2}{k} - \left(1 + \frac{4}{k} \right) e^{-k} \right)$$

$$\begin{cases} S_0 - 3S_2 = \frac{b}{k} (1 - e^{-k}) - 3 \frac{b}{k} \left(\frac{2}{k} - \left(1 + \frac{4}{k} \right) e^{-k} \right) = \frac{b}{k} \left(1 - \frac{6}{k} - \left(4 + \frac{12}{k} \right) e^{-k} \right) \\ S_0 - 2S_1 = \frac{b}{k} (1 - e^{-k}) - 2 \frac{b}{k} \left(\frac{1}{k} - \left(1 + \frac{1}{k} \right) e^{-k} \right) = \frac{b}{k} \left(1 - \frac{2}{k} - \left(3 + \frac{2}{k} \right) e^{-k} \right) \end{cases}$$

$$L = \frac{S_0 - 3S_2}{S_0 - 2S_1} = \frac{1 - \frac{6}{k} - \left(4 + \frac{12}{k} \right) e^{-k}}{1 - \frac{2}{k} - \left(3 + \frac{2}{k} \right) e^{-k}} = \frac{k - 6 + (4k + 12)e^{-k}}{k - 2 + (3k + 2)e^{-k}} \quad (14)$$

Значения $L(k)$ табулируются. После этого нахождение $k(L)$ не вызывает проблем. Далее находим

$$b = \frac{k(S_0 - 2S_1)}{1 - \frac{2}{k} - \left(3 + \frac{2}{k} \right) e^{-k}} \quad (15)$$

$$a = S_0 - \frac{b}{k} (1 - e^{-k}) \quad (16)$$

Модель Гомперца $y'/y = ae^{-kx}$; $y = y_\infty e^{a e^{-kx}}$.

$u = y'/y = ae^{-kx}$ представляет собой простую экспоненту, а ее параметры без дальнейшего дифференцирования получаются по приведенному выше методу.

Сигмоидальные модели

$$y = a + \frac{b(x-c)}{|x-c|+1}; \quad y = a + \frac{b(x-c)}{\sqrt{(x-c)^2+1}}; \quad y = a + \text{barctg}(k(x-c))$$

Находим точку центральной симметрией сигмоидальной модели $(c, y(c))$ y_c^* и переходим к $z = y - y_c^*$ – сигмоиде с нулем в точке c .

Для первого случая получим

$$z = \frac{b(x-c)}{k|x-c|+1}; \quad kz|x-c| - b(x-c) + z = 0 \quad (17)$$

Поскольку z и c известны, то методом наименьших квадратов находятся параметры b и k .

Для второго случая получим

$$z = \frac{b(x-c)}{\sqrt{k^2(x-c)^2+1}}; \quad z^2 = \frac{b^2(x-c)^2}{k^2(x-c)^2+1}; \quad |b^2 = b^*, k^2 = k^*|$$

$$k^* z^2 (x-c)^2 - b^* (x-c)^2 + z^2 (x-c)^2 = 0 \quad (18)$$

Поскольку z и c известны, то методом наименьших квадратов находятся параметры b^* и k^* и, следовательно b и k .

Для третьего случая получим

$$z = \text{barctg}(k(x-c)); z' = \frac{bk}{k^2(x-c)^2 + 1}; |bk = b^*, k^2 = k^*|$$

$$k^* z'(x-c)^2 - b^* + z' = 0 \quad (19)$$

Поскольку z , z' и c известны, то методом наименьших квадратов находятся параметры b^* и k^* и, следовательно b и k .

Нахождение технического тренда

Рассматриваемые выше методы выявления тренда основаны на проверке их соответствия исходным данным выдвинутым ранее некоторым гипотезам. Альтернативой являются чисто технические методы удаления колебаний (варианты скользящей средней). Пусть $y(t) = f(t) + g(t)$, где $g(x)$ описывает колебательный процесс с периодом T вокруг 0, а $f(t)$ – тренд, тогда можно ввести понятие T -тренда $u(t)$.

$$u(x) = \frac{1}{T} \int_{x-\frac{T}{2}}^{x+\frac{T}{2}} y(t) dt = \frac{1}{T} \int_{x-\frac{T}{2}}^{x+\frac{T}{2}} (f(t) + g(t)) dt = \frac{1}{T} \int_{x-\frac{T}{2}}^{x+\frac{T}{2}} f(t) dt \quad (20)$$

И принять функцию $u(x)$ за линию тренда. При этом необходимо помнить, что $u(t)$ дает систематическое отклонение от $f(t)$.

Представим

$$f(t) = f(x) + \sum_{k=1}^m \frac{f^{(k)}(x)}{k!} t^k,$$

Тогда

$$u(x) \approx \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left(f(x) + \sum_{k=1}^{2m} \frac{f^{(k)}(x)}{k!} t^k \right) dt = f(x) + \frac{1}{T} \sum_{k=1}^{2m} \frac{f^{(k)}(x)}{(k+1)!} t^{k+1} \Big|_{-T/2}^{T/2}$$

$$= f(x) + \frac{2}{T} \sum_{k=1}^m \frac{f^{(2k)}(x) T^{2k+1}}{(k+1)! 2^{2k+1}} = f(x) + \sum_{k=1}^m \frac{f^{(2k)}(x) T^{2k}}{(k+1)! 2^{2k}}$$

$$u(x) \approx f(x) + \frac{f''(x) T^2}{24} \quad (21)$$

Для оценки среднего значения $f''(t)$ введем функцию вида $ax^2 + bx + c$ аппроксимирующую исходную на интервале $x-T/2 \leq t \leq x+T/2$. Она имеет постоянную вторую производную, значение которой будем принимать за среднее значение $f''(t)$.

Ортогональные полиномы до второй степени на интервале $[-h, h]$ ($h=T/2$) имеют вид: $p_0=1$; $p_1=t$; $p_2=t^2-h^2/3$.

Коэффициент при второй степени аппроксимирующего полинома

$$a = \frac{\int_{-h}^h y(x+t) p_2(t) dt}{\int_{-h}^h p_2^2(t) dt} = f''(\theta)/2$$

$$\int_{-h}^h (t^2 - h^2/3)^2 dt = \frac{t^5}{5} - \frac{2t^3}{3} \frac{h^2}{3} + t \frac{h^4}{9} \Big|_{-h}^h = 2h^5 \left(\frac{1}{5} - \frac{1}{9} \right) = \frac{8h^5}{45} = \frac{T^5}{180}$$

$$f''(\theta) = \frac{360}{T^5} \int_{-T/2}^{T/2} y(x+t)p_2(t)dt$$

Соответственно, исправленное значение

$$f(x) \approx u(x) - \frac{f''(x)T^2}{24} = u(x) - \frac{15}{T^5} \int_{-T/2}^{T/2} y(x+t)p_2(t)dt \quad (22)$$

Нахождение колебаний

Колебания можно рассматривать, как разность между исходными данными и трендом. Анализ колебаний включает выявление их частотных и амплитудных составляющих, однако амплитуда колебаний может естественным образом зависеть от величины измеряемой характеристики.

Альтернативой для выявления колебательных процессов являются методы удаления тренда. В основе их лежит построение функции $u(f(t))$, которая поглощает трендовую составляющую $f(t)$ при сохранении ее колебательной составляющей [9, 10]. Примером такой функции может служить $u(t)=f(t+h)-2f(t)+f(t-h)$. В самом деле, пусть на некотором интервале $a \leq t \leq b$ представима в виде $f(t)=(pt+q)+g(x)$, где $g(x)$ описывает колебательный процесс вокруг 0. В этом случае $u(t)=f(t+h)-2f(t)+f(t-h)=g(t+h)-g(t)+g(t-h)$ и $u(t)$ не содержит трендовой составляющей, но сохраняет колебательные составляющие. Еще один вариант функции $u(t)=f(t+h)f(t-h)/f(t)^2$ или $u(t) = \ln \left(\frac{f(t+h)f(t-h)}{f(t)^2} \right)$, обеспечивающий удаление экспоненциального тренда. Отметим, что последний вариант сводится к линейному переходом в полулогарифмические координаты $F(t)=\ln(f(t))$.

Определение периодов колебаний функции $u(t)$ может быть выполнено, например, на основе быстрого преобразования Фурье или с использованием сдвиговой функции Альтера – Джонсона.

К сожалению, функции $u(t)$ заметно искажают амплитуды колебательных процессов. Пусть $f(t)=(pt+q)+g(x)=(pt+q)+A\sin(2\pi t/T+\varphi)$. В этом случае

$$\begin{aligned} u(t) &= A \left(\sin \left(\frac{2\pi}{T}t + \varphi + \frac{2\pi}{T}h \right) + \sin \left(\frac{2\pi}{T}t + \varphi - \frac{2\pi}{T}h \right) - 2\sin \left(\frac{2\pi}{T}t + \varphi \right) \right) \\ &= A \left(2\sin \left(\frac{2\pi}{T}t + \varphi \right) \cos \left(\frac{2\pi}{T}h \right) - 2\sin \left(\frac{2\pi}{T}t + \varphi \right) \right) \\ &= 2A \left(\cos \left(\frac{2\pi}{T}h \right) - 1 \right) \sin \left(\frac{2\pi}{T}t + \varphi \right) = 4A \sin^2 \left(\frac{\pi h}{T} \right) \sin \left(\frac{2\pi}{T}t + \varphi \right) \\ u(t) &= 4A \sin^2 \left(\frac{\pi h}{T} \right) \sin \left(\frac{2\pi}{T}t + \varphi \right) \end{aligned} \quad (23)$$

Таким образом амплитуда колебаний $B(T)=4A\sin^2(\pi h/T)$ при $T \ll h$ превращается в практически генератор случайных чисел, при $T \gg h$ стремится к 0. В этих условиях достоверное определение периодов колебаний возможно только для ограниченного диапазона длин этих периодов.

Пусть исходный процесс описывается функцией $y(t)$, тренд – функцией $U(t)$, тогда

колебательная составляющая представима в виде $\Phi(t)=y(t)-U(t)$. Для оценки тренда амплитуды можно воспользоваться функцией $Z(t)=|\Phi(t)|$, усреднение которой дает величину, пропорциональную амплитуде (или $Z(t)=\Phi^2(t)$ пропорциональную квадрату амплитуды). Для получения собственно линии тренда можно воспользоваться предварительным ее усреднением по периоду T колебаний $Z(t)$, либо по представлению $R(t)=\frac{1}{T}\int_{x-T/2}^{x+T/2} Z(x)dx$, либо используя алгоритм удаления шума, описанный выше, где за величину шума $s(t)$ в точке t принимается $s(t)=Z(t)-R(t)$. В последнем случае не происходит явная потеря концов интервала, хотя точность на концах, конечно, получается невысокой. Полученное выражение $A(t)$ пригодно для аналитической аппроксимации либо непосредственно, либо как отношение $A(t)/U(t)$, либо как сумма вида $a+bU(t)$, где a и b некоторые константы, а $U(t)$ – тренд характеристики. После этого колебание $\Phi(t)$ представимо в виде $\Phi(t)=A_T(t)\varphi(t)$.

Здесь $A_T(t)$ – тренд амплитуды, а $\varphi(t)$ представляет колебания с почти постоянной амплитудой и может быть относительно успешно представлена аналитическим выражением.

Функция, описывающая развитие процесса во времени роста (темпов развития) в принятых обозначениях представима в виде

$$y = U(t) + A_T(t)(a + bU(t))\varphi(t) \quad (24)$$

или для темпов

$$\frac{y'}{y} = U(t) + (a + bU(t))\varphi(t)$$

$$y = y_{\infty} e^{U(t) + (a + bU(t))\varphi(t)} \quad (25)$$

Анализ колебательной составляющей $\varphi(t)$ можно проводить на основе ее представления в виде суммы нескольких простых периодических колебаний. Периоды этих колебаний можно определить на основе применения дискретного преобразования Фурье (ДПФ), нахождения минимумов сдвиговой функции или простым подбором гармонических колебаний.

Для сдвиговой функции (функция Джонсона) [9] $p(\tau)$ определяются ее локальные минимумы, соответствующие предполагаемым почти периодам в исходном временном ряде. Значения $p(\tau)$ определяются из

$$p(\tau) = \frac{1}{n-\tau} \frac{1}{y_{\max} - y_{\min}} \sum_{i=1}^{n-\tau} |y_{i+\tau} - y_i| \quad (26)$$

Подбор гармонических колебаний осуществляется минимизацией функции $R(T_1, T_2, \dots)$ по параметрам T_1, T_2, \dots . $R(T_1, T_2, \dots)$ из минимизации

$$R(T_1, T_2, \dots) = \min_{a_k, b_k} \sum_{k=0}^n \left(\varphi(x_k) - \sum_k \left(a_k \sin\left(\frac{2\pi}{T_k} x_k\right) + b_k \sin\left(\frac{2\pi}{T_k} x_k\right) \right) \right)^2 \quad (27)$$

В ряде случаев, например при анализе темпов развития процесса, то есть величин вида y'/y с несимметричными колебаниями вокруг 0 прямое нахождение тренда, отличного от линейного может оказаться затруднительным. В этом случае оказывается целесообразным предварительное определение технического тренда с усреднением по периоду колебаний, а затем построение уже на нем аналитического тренда в

соответствии с выдвинутой гипотезой.

Численная реализация

Рассмотрим предлагаемые методы на примере – статистике производства товаров длительного пользования в США за 1947-2016 годы (Bureau of Economic Analysis, United States Department of Commerce. https://www.bea.gov/sites/default/files/2018-04/GDPbyInd_VA_1947-2017.xlsx). Соответствующие данные представлены на рисунке 1. Полученная удалением шума кривая является гладкой и пригодна для обработки с использованием дифференциальных соотношений.

На рисунке 2 приведены 3 кривые, описывающие фактические темпы развития производства, темпы с удалением периодической составляющей (с почти периодом в 10,5 лет) и убывающим экспоненциальным трендом. Сразу заметим, что при переходе от абсолютных величин к темпам явно видна циклическая составляющая с почти периодом $\approx 10,5$ лет (циклы Жигляра – Маркса). Удаление этой составляющей показывает наличие и более длинных циклов (похожих на Кондратьевские, хотя и более коротких). Тренд близок к убывающему экспоненциальному. Последнее говорит, что процесс развивается по схеме близкой к «испорченной» модели Гомперца, где в экспоненту включается колебательная составляющая.



Рисунок 1. Динамика производства товаров длительного пользования в США за 1947-2016 годы (фактическая и с удалением шума).



Рисунок 2. Динамика темпов развития производства с выделением среднесрочного тренда с удалением коротких колебаний и долговременного аperiодического экспоненциального тренда.

На рисунке 3 представлены абсолютные и относительные колебания темпов относительно аperiодического тренда и результаты попыток их аппроксимации суммой 2 и 3 гармонических колебаний.

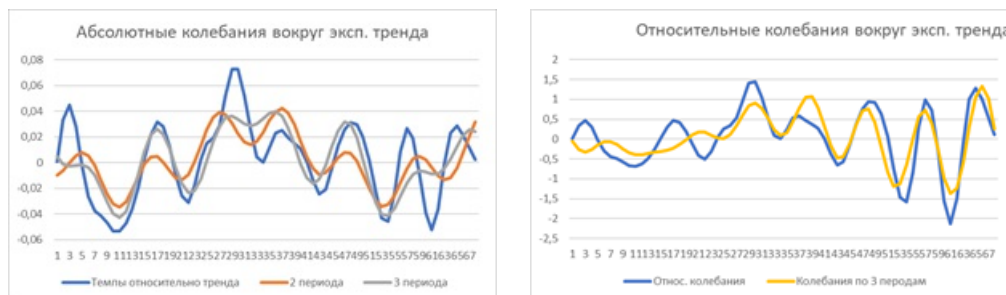


Рисунок 3. Пример разложения абсолютных и относительных колебаний темпов относительно аperiодического тренда в виде суммы гармоник.

Аппроксимация, «похожа» на исходные колебания, но не более того. Это отражает тот факт, что колебания являются именно почти периодическими и строгих периодов не имеют. Последнее принципиально ограничивает возможности долговременного прогноза в силу неопределенности конкретных сроков наступления тех или иных событий и соответствующих им значений, сохраняя лишь качественный ход процесса (почти как у Нострадамуса), давая почву для «оптимистических» и «пессимистических» прогнозов.

Заключение

Рассмотренные методы позволяют проводить достаточно объективный анализ данных и создают условия для построения обоснованного прогноза. Проведен численный анализ по данным многолетней статистики динамики производства. Научная новизна исследования заключается в разработке методики декомпозиции процесса на трендовые и колебательные компоненты. В отличие от большинства существующих исследований анализа динамики процессов большое внимание уделяется учету и оценке уровня шума с определением пределов точности получаемых результатов и, тем более, прогнозов, что позволяет избегать необоснованных выводов и решений и построения «слишком» точных результатов на основе недостаточно точных исходных данных, исходя из требований гладкости функции при имеющемся уровне шума. Использование функций ограниченного роста и выявление точек смены тренда позволяет проводить корректное качественное долгосрочное прогнозирование без необоснованного предсказания катастрофического

хода изучаемого процесса. Полученные результаты позволяют получить аналитическое выражение изучаемых процессов на основе статистических данных, представленных в виде временных рядов что позволяет не только аппроксимировать поведение процесса, но и выявлять его физическую сущность и, соответственно использовать полученные решения на целый класс процессов аналогичной природы. Предлагаемые методы и алгоритмы ориентированы, прежде всего на анализ и прогнозирование динамики экономических процессов, а также ряда природных процессов, например, метеорологических, подверженных влиянию различных, в том числе и случайных факторов.

Библиография

1. Breiman L., Friedman J., Olshen R.A., Stone C.J. Classification and Regression Trees. New York: Imprint Chapman and Hall/CRC, 2017. – 368 p. DOI: <https://doi.org/10.1201/9781315139470>. (First published 1984)
2. Дюран Б., Оделл П. Кластерный анализ. М.: Статистика, 1977. – 128 с.
3. Рашид Тарик. Создаем нейронную сеть. М.: Диалектика-Вильямс, 2023. – 272 с.
4. Мошков М. Ю. Оценки глубины деревьев решений над конечными двужначными системами проверок. // Математические вопросы кибернетики. Вып. 7. М.: Физматлит, 1998. С. 161-168. URL: <http://library.keldysh.ru/mvk.asp?id=1998-161>.
5. Savitzky A., Golay M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry. 1964. 36(8). P. 1627-1639.
6. Скляр А.Я. Анализ и устранение шумовой компоненты во временных рядах с переменным шагом // Кибернетика и программирование. 2019. № 1. С. 51-59. DOI: 10.25136/2644-5522.2019.1.27031 URL: https://nbpublish.com/library_read_article.php?id=27031
7. Скляр А. Я. Математическое моделирование экономических процессов на основе принципа максимума полезности. М.: РТУ МИРЭА, 2021. – 180 с.
8. Кузьмин В.И., Самохин А.Б., Гадзаов А.Ф., Чердынцев В.В. Модели и методы определения параметров нелинейных процессов. М.: Московский технологический университет (МИРЭА), 2016. – 148 с.
9. Johnson M. Correlations of cycles in weather, solar activity, geomagnetic values and planetary configurations. San Francisco: Phillips and Van Orden, 1944. – 122 p.
10. Кузьмин В.И., Самохин А.Б. Почти периодические функции с трендом. Вестник МГТУ МИРЭА. 2015. № 4(9). Т. 2. С. 105-107. EDN: VNIYPJ

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Предметом исследования в рецензируемом исследовании выступают методы анализа временных рядов, в статье акцент сделан на нелинейные регрессионные модели. Методология работы базируется на применении различных подходов для аппроксимации трендов, статистического сглаживания (полиномиальные фильтры, оценка уровня шума через дисперсию), гармонического анализа для выделения периодических компонент, а также визуализации данных.

Актуальность исследования авторы связывают с необходимостью анализа статистических данных для выявления внутренних зависимостей.

Научная новизна исследования заключается в обобщении сведений о моделях и алгоритмах нелинейного регрессионного анализа временных рядов; полученное в

результате исследования приращение научного знаниями позиционируется авторами как разработка «методики декомпозиции процесса на трендовые и колебательные компоненты».

Структурно в тексте выделены следующие озаглавленные разделы и подразделы: Введение, Оценка и удаление шума в данных, Нахождение долговременного тренда, Дифференциальные методы (Экспоненциальное приближение к постоянной, Гиперболическое приближение к постоянной, Модель Гомперца, Сигмоидальные модели), Интегральные методы, Характеристические показатели (Экспоненциальное приближение к постоянной), Нахождение технического тренда, Нахождение колебаний, Численная реализация, Заключение и Библиография.

В статье рассмотрены такие этапы анализа временных рядов как удаление шума (методы Савицкого–Галея, скользящие средние), выделение тренда (экспоненциальные, гиперболические, сигмоидальные модели), анализ колебательных компонент (разложение Фурье, сдвиговые функции). Авторами освещены как классические подходы (метод наименьших квадратов), так и оригинальные алгоритмы, например, интегральные методы для оценки параметров тренда. Численная реализация рассмотренных в статье методов проиллюстрирована авторами на примере – статистике производства товаров длительного пользования в США за 1947-2016 годы. Предлагаемые методы и алгоритмы ориентированы на анализ и прогнозирование динамики экономических процессов, а также ряда природных процессов, например, метеорологических, подверженных влиянию различных, в том числе и случайных факторов.

Библиографический список включает 10 источников – научные публикации российских и зарубежных авторов по рассматриваемой теме на русском языке и иностранных языках. В тексте публикации имеются адресные отсылки к списку литературы, подтверждающие наличие апелляции к оппонентам.

Из резервов улучшения работы стоит отметить, что авторами не выполнены требования принятых редакцией Правил оформления списка литературы, в соответствии с которыми «рекомендованный объем списка литературы для оригинальной научной статьи – не менее 20 источников». Оформление интервалов, которые в представленных материалах отражаются в виде отсылок к списку литературы ввиду использования квадратных скобок, например «к интервалу $[0, 1]$ » (с использованием для написания " $[0, 1]$ " в качестве шрифта верхнего индекса).

Тема статьи актуальна, материал соответствует тематике журнала «Программные системы и вычислительные методы», отражает результаты проведенного исследования, может вызвать интерес у читателей, рекомендуется к опубликованию.