

УДК 519.2:303.732.4

UDC 519.2:303.732.4

**ВОССТАНОВЛЕНИЕ ЗАВИСИМОСТИ
МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ НА
ОСНОВЕ НЕПАРАМЕТРИЧЕСКОЙ МОДЕЛИ
С ПЕРИОДИЧЕСКОЙ СОСТАВЛЯЮЩЕЙ****RENEWAL DEPENDENCE METHOD OF
LEAST SQUARES BASED NONPARAMETRIC
MODEL WITH PERIODIC COMPONENT**

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci., professor

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

Bauman Moscow State Technical University, Moscow, Russia

Рассмотрена непараметрическая задача восстановления зависимости, которая описывается суммой линейного тренда и периодической функции с известным периодом. Получены асимптотические распределения оценок параметров и трендовой составляющей. Разработаны методы оценивания периодической компоненты и построения интервального прогноза. В рамках модели точек наблюдения, естественной для приложений, обоснованы условия применимости. В частности, установлена асимптотическая несмещенность оценки коэффициента линейного члена

We consider the nonparametric problem of renewal dependence, which is described by the sum of a linear trend and periodic function with a known period. We obtain the asymptotic distribution of the parameter estimates and the trend component. The methods of estimating the periodic component and designing interval forecast. In the model of the points of observation, natural for applications, justified by the conditions of use. In particular, we prove an asymptotically unbiased estimate of the coefficient of the linear term

Ключевые слова: МЕТОД НАИМЕНЬШИХ
КВАДРАТОВ, НЕПАРАМЕТРИЧЕСКИЕ
МЕТОДЫ, ПЕРИОДИЧЕСКАЯ
СОСТАВЛЯЮЩАЯ, ОЦЕНИВАНИЕ,
ПРОГНОЗИРОВАНИЕ

Keywords: METHOD OF LEAST SQUARES,
NONPARAMETRIC METHODS, PERIODIC
COMPONENT, ESTIMATION, FORECASTING

Метод наименьших квадратов восстановления зависимости – один из наиболее распространенных статистических методов анализа данных. В статье рассмотрена непараметрическая постановка: восстанавливаемая зависимость – сумма линейной функции и периодической составляющей произвольного вида (с известным периодом), распределение случайных погрешностей (остатков, невязок) произвольно.

1. Задача восстановления линейной зависимости

Начнем с простейшего случая – задачи восстановления линейной зависимости. Пусть t – независимая переменная, а x – зависимая. Рассмотрим задачу восстановления зависимости $x = x(t)$ на основе набора n пар чисел (t_k, x_k) , $k = 1, 2, \dots, n$, где t_k – значения независимой переменной, а x_k – соответствующие им значения зависимой переменной.

Восстанавливать зависимость можно на основе различных моделей. Обычно применяют модели временных рядов, включающие три составляющие: трендовую (T), периодическую (S) и случайную (E). Рассматривают, как в [1] и аналогичных изданиях, аддитивную модель $T + S + E$ и мультипликативную модель $T \times S \times E$.

Простейшая аддитивная модель имеет вид

$$x_k = a(t_k - \bar{t}) + d + e_k = a(t_k - \bar{t}) + d + f(t_k) + E_k, \quad k = 1, 2, \dots, n. \quad (1)$$

Здесь трендовая составляющая – линейная функция $a(t_k - \bar{t}) + d$ (такая запись тренда предпочтительнее для облегчения выкладок); периодическая составляющая $f(t)$ обычно описывает сезонность, т.е. период известен (в зависимости от моделируемой ситуации он равен году, неделе, суткам и т.п.); случайная составляющая представлена слагаемыми E_k , которые являются реализациями независимых одинаково распределенных случайных величин с нулевым математическим ожиданием и дисперсией s^2 , неизвестной статистике. В рассматриваемой модели $e_k = f(t_k) + E_k, k = 1, 2, \dots, n$, т.е. отклонения от линейного тренда e_k не являются одинаково распределенными. Однако их распределения отличаются лишь сдвигами (на значения детерминированной периодической составляющей).

Соответствующая модели (1) мультипликативная модель имеет вид

$$y_k = [Bt_k^a] \times f_1(t_k) \times [1 + e_k], \quad k = 1, 2, \dots, n. \quad (2)$$

В модели (2) сомножители имеют описанный выше смысл. При логарифмировании модель (2) переходит в аналог модели (1), следовательно, достаточно рассматривать модель (1).

Иногда принимают предположение о нормальности распределения погрешностей. Однако давно известно, что распределения реальных данных, как правило, отличаются от нормальных [2]. Поэтому далее рассматриваем *непараметрическую модель*, не предполагающую, что распределение погрешностей входит в то или иное параметрическое семейство. Отказ от задания распределения погрешностей в параметрическом виде – одно из

оснований для того, чтобы именовать рассматриваемые модель и метод непараметрическими. Второе основание – отказ от выбора периодической составляющей из какого-либо параметрического семейства функций.

Практическая значимость модели (1) очевидна. Однако расчетные методы, описанные в [1] и аналогичных изданиях, являются эвристическими. Цель настоящей статьи – *построить непараметрическую вероятностно-статистическую теорию прогноза временного ряда на базе линейного тренда с учетом аддитивной периодической составляющей*.

Метод наименьших квадратов был разработан К. Гауссом в 1794 г. [2]. Согласно нему этому методу для расчета наилучшей функции, приближающей линейным образом зависимость x от t в модели (1), следует рассмотреть функцию двух переменных

$$f(a, d) = \sum_{k=1}^n (x_k - a(t_k - \bar{t}) - d)^2.$$

Оценки метода наименьших квадратов (кратко: оценки МНК) – это такие значения a^* и d^* , при которых функция $f(a, d)$ достигает минимума по всем значениям аргументов. Как известно (см., например, [2]), оценки МНК имеют вид

$$a^* = \frac{\sum_{k=1}^n x_k (t_k - \bar{t})}{\sum_{k=1}^n (t_k - \bar{t})^2}, \quad d^* = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k. \quad (3)$$

Следуя эвристическому подходу [1], изучим асимптотическое поведение оценок МНК a^* и d^* , заданных формулами (3), установим их асимптотическую нормальность в предположениях модели (19), а затем состоятельно оценим периодическую составляющую $f(t)$ и построим интервальный прогноз для $x(t)$.

2. Асимптотические распределения оценок параметров

Из формулы (3) следует, что

$$d^* = \frac{a}{n} \sum_{k=1}^n (t_k - \bar{t}) + d + \frac{1}{n} \sum_{k=1}^n e_k = d + \frac{1}{n} \sum_{k=1}^n e_k = d + \frac{1}{n} \sum_{k=1}^n f(t_k) + \frac{1}{n} \sum_{k=1}^n E_k. \quad (4)$$

Согласно Центральной предельной теореме (для выполнения ее условий необходимо предположить, например, что погрешности e_k , $k = 1, 2, \dots, n$, финитны или имеют конечный третий абсолютный момент; однако заострять внимание на этих внутриматематических «условиях регулярности» здесь нет необходимости) оценка d^* имеет асимптотически нормальное распределение с математическим ожиданием $d + \frac{1}{n} \sum_{k=1}^n f(t_k)$ и дисперсией s^2/n , ее оценка приводится ниже. Из формул (3) и (4) вытекает, что

$$x_k - \bar{x} = a(t_k - \bar{t}) + d + e_k - d - \frac{1}{n} \sum_{k=1}^n e_k = a(t_k - \bar{t}) + e_k - \frac{1}{n} \sum_{k=1}^n e_k,$$

$$(x_k - \bar{x})(t_k - \bar{t}) = a(t_k - \bar{t})^2 + e_k(t_k - \bar{t}) - \frac{(t_k - \bar{t})}{n} \sum_{k=1}^n e_k.$$

Последнее слагаемое во втором соотношении при суммировании по k обращается в 0, поэтому

$$a^* = a + \sum_{k=1}^n c_k e_k = a + \sum_{k=1}^n c_k f(t_k) + \sum_{k=1}^n c_k E_k, \quad c_k = \frac{(t_k - \bar{t})}{\sum_{k=1}^n (t_k - \bar{t})^2}. \quad (5)$$

Формулы (5) показывают, что оценка a^* является асимптотически нормальной с математическим ожиданием $a + \sum_{k=1}^n c_k f(t_k)$ и дисперсией

$$D(a^*) = \sum_{k=1}^n c_k^2 D(E_k) = \frac{s^2}{\sum_{k=1}^n (t_k - \bar{t})^2}.$$

Отметим, что многомерная нормальность имеет быть, когда каждое слагаемое в формуле (5) мало сравнительно со всей суммой, т.е.

$$\lim_{n \rightarrow \infty} \max |t_k - \bar{t}| / \left\{ \sum_{k=1}^n (t_k - \bar{t})^2 \right\}^{1/2} = 0. \quad (6)$$

Условие (6) выполнено, например, если t_k образуют (полную, т.е. без пропусков) арифметическую прогрессию, число членов которой безгранично растёт.

Итак, дисперсии оценок МНК параметров a^* и d^* линейного тренда – те же, что и при отсутствии сезонных искажений (см., например, [2]). А вот их математические ожидания зависят от периодической составляющей. Однако в случае

$$\sum_{i=1}^n f(t_i) = 0, \quad \sum_{i=1}^n (t_i - \bar{t})f(t_i) = 0 \quad (7)$$

оценки a^* и d^* являются несмещёнными.

Условия (7) являются необходимыми и достаточными для несмещённости и состоятельности оценок МНК коэффициентов линейной зависимости. Проверка условий (7) рассмотрена в конце статьи.

Несмещённость (в предположениях (7)) и асимптотическая нормальность оценок метода наименьших квадратов позволяют легко указывать для них асимптотические доверительные границы и проверять статистические гипотезы, например, о равенстве определённым значениям, прежде всего 0.

3. Асимптотическое распределение трендовой составляющей

Из формул (4) и (5) следует, что при справедливости соотношений (7)

$$M\{a^*(t - \bar{t}) + d^*\} = M(a^*)(t - \bar{t}) + M(d^*) = a(t - \bar{t}) + d,$$

т.е. оценка $y^*(t) = a^*(t - \bar{t}) + d^*$ трендовой составляющей $y(t) = a(t - \bar{t}) + d$ рассматриваемой зависимости является несмещённой. Поэтому

$$D(y^*(t)) = D(a^*)(t - \bar{t})^2 + 2M\{(a^* - a)(d^* - d)(t - \bar{t})\} + D(d^*).$$

При этом, поскольку погрешности E_k независимы в совокупности и $M(E_k) = 0$, то

$$M\{(a^* - a)(d^* - d)(t - \bar{t})\} = \frac{1}{n} \sum_{k=1}^n c_k (t - \bar{t}) M(E_k^2) = \frac{1}{n} (t - \bar{t}) S^2 \sum_{k=1}^n c_k = 0.$$

Таким образом,

$$D(y^*(t)) = s^2 \left\{ \frac{1}{n} + \frac{(t-\bar{t})^2}{\sum_{k=1}^n (t_k - \bar{t})^2} \right\}. \quad (8)$$

Итак, оценка $y^*(t)$ является несмещенной и асимптотически нормальной. Для ее практического использования (построения доверительных интервалов, проверки статистических гипотез) необходимо состоятельно уметь оценивать остаточную дисперсию $M(E_k^2) = s^2$.

В частности, не представляет труда выписывание нижней и верхней границ для трендовой составляющей прогностической функции:

$$y_{нижн}(t) = a^*(t-\bar{t}) + d^* - d(t), \quad y_{верх}(t) = a^*(t-\bar{t}) + d^* + d(t),$$

где полуширина доверительного интервала $d(t)$ имеет вид

$$d(t) = U(g) \sqrt{D^*(y^*(t))} = U(g) s^* \sqrt{\frac{1}{n} + \frac{(t-\bar{t})^2}{\sum_{k=1}^n (t_k - \bar{t})^2}}. \quad (9)$$

Здесь g - доверительная вероятность, $U(g)$ - квантиль нормального распределения порядка $\frac{1+g}{2}$, т.е. $U(g) = \Phi^{-1}\left(\frac{1+g}{2}\right)$, где $\Phi(x)$ - функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1. При $g = 0,95$ (наиболее применяемое значение) имеем $U(g) = 1,96$. В формуле (9) $D^*(y^*(t))$ - состоятельная оценка дисперсии $y^*(t)$. В соответствии с (8) она является произведением состоятельной оценки s^* среднего квадратического отклонения s случайных погрешностей E_k на известную исследователю детерминированную функцию от t .

4. Математическое ожидание остаточной суммы квадратов

В точках $t_k, k = 1, 2, \dots, n$, имеются исходные значения зависимой переменной x_k и восстановленные значения $y^*(t_k)$. Рассмотрим остаточную сумму квадратов

$$SS = \sum_{k=1}^n (y^*(t_k) - x_k)^2 = \sum_{k=1}^n \{(a^* - a)(t_k - \bar{t}) + (d^* - d) - f(t_k) - E_k\}^2.$$

При отсутствии периодической составляющей используют [2] состоятельные оценки s^* среднего квадратического отклонения s случайных погрешностей, построенные на основе остаточной суммы квадратов $s^* = \sqrt{\frac{SS}{n}}$ или $s^* = \sqrt{\frac{SS}{n-2}}$. Однако при наличии периодической составляющей так делать нельзя. Приходится использовать «обходный путь».

В соответствии с формулами (4) и (5) при справедливости условий (7)

$$\begin{aligned} SS &= \sum_{k=1}^n \left\{ (t_k - \bar{t}) \sum_{j=1}^n c_j E_j + \frac{1}{n} \sum_{j=1}^n E_j - f(t_k) - E_k \right\}^2 = \\ &= \sum_{k=1}^n \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j - f(t_k) - E_k \right\}^2 = \sum_{k=1}^n SS_k. \end{aligned}$$

Найдем математическое ожидание каждого из слагаемых:

$$\begin{aligned} M(SS_k) &= M \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j - f(t_k) - E_k \right\}^2 = M \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j \right\}^2 - \\ &- 2M \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j \right\} (f(t_k) + E_k) + M(f(t_k) - E_k)^2. \end{aligned}$$

Поскольку E_k независимы, одинаково распределены и имеют нулевое математическое ожидание, то

$$M \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j \right\}^2 = \sum_{j=1}^n \left\{ c_j (t_k - \bar{t}) + \frac{1}{n} \right\}^2 s^2.$$

Далее,

$$-2M \left\{ \sum_{j=1}^n \left[c_j (t_k - \bar{t}) + \frac{1}{n} \right] E_j \right\} (f(t_k) + E_k) = -2 \left\{ c_k (t_k - \bar{t}) + \frac{1}{n} \right\} s^2.$$

Наконец,

$$M(f(t_k) - E_k)^2 = f^2(t_k) + s^2.$$

На основе трех последних равенств можно показать, что при выполнении условия асимптотической нормальности (6)

$$\lim_{n \rightarrow \infty} M(SS_k) = f^2(t_k) + s^2.$$

Следовательно,

$$M\left(\frac{SS}{n}\right) = s^2 + \frac{1}{n} \sum_{k=1}^n f^2(t_k). \quad (10)$$

В правой части (10) первое слагаемое соответствует вкладу случайной составляющей, второе – вкладу периодической составляющей.

В некоторых случаях второе слагаемое в правой части (10) может быть известно из предыдущего опыта или же оценено экспертами, однако в большинстве ситуаций целесообразно исходить из оценки периодической составляющей.

5. *Оценивание периодической составляющей*

В литературе рассматривают как параметрические, так и непараметрические подходы. Популярный метод исходит из того, что достаточно гладкую функцию можно разложить в ряд Фурье и получить хорошее приближение с помощью небольшого числа гармоник. В простейшем случае – одна гармоника. Так, динамику индекса инфляции можно попытаться изучать с помощью модели

$$x_k = a(t_k - \bar{t}) + d + f(t_k) + E_k = a(t_k - \bar{t}) + d + g \cos(2\pi t_k) + E_k, \quad k = 1, 2, \dots, n$$

(время t измеряется в годах). Тогда неизвестные параметры a , b , g оцениваются методом наименьших квадратов.

Однако обычно нет оснований предполагать, что периодическая составляющая входит в то или иное параметрическое семейство функций. Приходится строить непараметрические оценки. Опишем одну из возможных постановок.

Пусть в согласии с предположениями (7) рассматривается целое число периодов, т.е. $n = tq$, где n – объем наблюдений, t – количество периодов, q – число наблюдений в одном периоде. Предполагается, что первые q моментов наблюдения при сдвиге на длину периода дают следующие q мо-

ментов времени, при сдвиге на две длины периода дают третий набор из q моментов наблюдения, и т.д. Тогда в соответствии с определением периодической составляющей справедливо равенство

$$f(t_s) = f(t_{q+s}) = f(t_{2q+s}) = \dots = f(t_{(m-1)q+s}), \quad s = 1, 2, \dots, q. \quad (11)$$

Если наблюдения проводятся ежемесячно в течение m лет, то число наблюдений в одном периоде $q = 12$, общий объем наблюдений $n = 12m$, далее s – номер месяца в году, $s = 1, 2, \dots, 12$. Пусть g_s – общее значение в (11). Для оценки периодической составляющей требуется оценить g_1, g_2, \dots, g_q .

Естественный подход состоит в том, чтобы усреднить m значений $x_k - y^*(t_k)$, соответствующих моментам времени, отстоящим друг от друга на целое число периодов. Другими словами, усреднить «очищенные» от трендовой составляющей исходные данные, соответствующие одноименным месяцам различных лет. Речь идет об оценках

$$g_s^* = \frac{1}{m} \sum_{j=1}^m (x_{s+(j-1)q} - y^*(t_{s+(j-1)q})), \quad s = 1, 2, \dots, q. \quad (12)$$

Оценка периодической составляющей распространяется на весь интервал наблюдений очевидным образом:

$$f^*(t_s) = f^*(t_{q+s}) = f^*(t_{2q+s}) = \dots = f^*(t_{(m-1)q+s}) = g_s^*, \quad s = 1, 2, \dots, q. \quad (13)$$

Сложив восстановленные значения трендовой и периодической составляющей, получим оценку зависимости, «очищенную» от случайной составляющей

$$x^*(t) = y^*(t) + f^*(t) = a^*(t - \bar{t}) + d^* + f^*(t). \quad (14)$$

Здесь оценки a^* и d^* находят по формулам (3), а оценки $f^*(t)$ – по формулам (12) – (13).

С помощью формулы (14) можно строить точечный прогноз, используя ее вне интервала наблюдений. Для этого достаточно распространить сезонную составляющую $f^*(t)$ вплоть до рассматриваемого момента времени по правилу (13) и суммировать ее с прогнозом трендовой составляющей

$y^*(t)$. Интерполяция и экстраполяция на моменты времени t , не входящие в исходное множество $\{t_k, k = 1, 2, \dots, n\}$ и множества, полученные из него сдвигами на целое число периодов, может быть осуществлена путем линейной интерполяции ближайших значений или иным методом сглаживания.

Обсудим свойства оценок (12) – (14).

При безграничном росте объема данных и справедливости условий (6) и (7) оценки a^* и d^* параметров трендовой составляющей являются состоятельными и несмещенными, а потому, как можно показать, в рассматриваемых в настоящей статье условиях суммы (12) оценивают периодическую составляющую состоятельно (при $m \rightarrow \infty$) и несмещенно. Как следствие,

$$\frac{1}{n} \sum_{k=1}^n [f^*(t_k)]^2 - \frac{1}{n} \sum_{k=1}^n f^2(t_k) \rightarrow 0 \quad (15)$$

по вероятности при $n \rightarrow \infty$. В соответствии с (10) последнее соотношение дает возможность оценить s^2 , а затем построить интервальный прогноз для трендовой составляющей согласно (9).

Отметим, что в рассматриваемой ситуации, как правило, n растет, увеличиваясь на величины, кратные q – числу наблюдений в одном периоде. Как следствие, уменьшаемое в (15) – константа, зависимости от n нет. Эти особенности связаны с тем, что выполнение условий (7) предполагает рассмотрение целого числа периодов.

Рассмотрим оценки (12) подробнее. Как вытекает из (4.1.19), (11) и (12),

$$g_s^* = f(t_s) - (a^* - a) \frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t}) - (d^* - d) + \frac{1}{m} \sum_{j=1}^m E_{s+(j-1)q}, \quad s = 1, 2, \dots, q.$$

С учетом (4), (5) и (7) получаем, что

$$g_s^* = f(t_s) - \left(\sum_{k=1}^n c_k E_k \right) \left(\frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t}) \right) - \frac{1}{n} \sum_{k=1}^n E_k + \frac{1}{m} \sum_{j=1}^m E_{s+(j-1)q}, \quad s = 1, 2, \dots, q.$$

Таким образом,

$$g_s^* = f(t_s) + \sum_{k=1}^n h_{ks} E_k, \quad s = 1, 2, \dots, q \quad (16)$$

где $h_{ks} = -c_k r_s - \frac{1}{n} + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $h_{ks} = -c_k r_s - \frac{1}{n}$ при всех остальных значениях индекса суммирования k , и $r_s = \frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t})$.

Соотношение (16) означает, что рассматриваемые оценки есть суммы независимых случайных величин, а потому с помощью Центральной предельной теоремы можно построить доверительные интервалы для рассматриваемых значений периодической составляющей (в предположении справедливости условий (6)).

6. Интервальный прогноз

Точечный прогноз строят по формуле (11) на основе $x^*(t)$ - оценки зависимости, «очищенной» от случайной составляющей, но включающей трендовый и периодический компоненты. Если выполнены условия (7), то

$$Mx^*(t) = x(t) = a(t - \bar{t}) + d + f(t),$$

т.е. оценка $x^*(t)$ является несмещенной.

При справедливости условий (7) с учетом (4), (5) и (16) получаем, что для момента времени t , входящего в исходное множество $\{t_k, k = 1, 2, \dots, n\}$ или в множества, полученные из него сдвигами на целое число периодов,

$$x^*(t) - x(t) = (t - \bar{t}) \sum_{k=1}^n c_k E_k + \frac{1}{n} \sum_{k=1}^n E_k + \sum_{k=1}^n h_{ks} E_k. \quad (17)$$

В (17) при определении значений коэффициентов h_{ks} в качестве s следует взять номер наименьшего из исходных моментов времени $\{t_k, k = 1, 2, \dots, n\}$, отстоящих от рассматриваемого момента t на целое число периодов. С помощью (16) заключаем, что

$$x^*(t) - x(t) = \sum_{k=1}^n w_{ks} E_k,$$

где $w_{ks} = c_k(t - \bar{t} - r_s) + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $w_{ks} = c_k(t - \bar{t} - r_s)$ при всех остальных значениях индекса суммирования k , и r_s – то же, что и в формуле (16).

В правой части формулы (17) стоит сумма независимых случайных величин, поэтому оценка $x^*(t)$ является асимптотически нормальной (при справедливости условий (6)) с математическим ожиданием $x(t)$ и дисперсией

$$D(x(t)) = \sum_{k=1}^n w_{ks}^2 D(E_k) = s^2 \sum_{k=1}^n w_{ks}^2. \quad (18)$$

Следовательно, нижняя $x_{\text{нижн}}(t)$ и верхняя $x_{\text{верх}}(t)$ доверительные границы для прогностической функции (с учетом как трендовой, так и периодической составляющих) имеют вид:

$$x_{\text{нижн}}(t) = a^*(t - \bar{t}) + d^* + f^*(t) - \Delta(t), \quad x_{\text{верх}}(t) = a^*(t - \bar{t}) + d^* + f^*(t) + \Delta(t),$$

где

$$\Delta(t) = U(g) \sqrt{D^*(x^*(t))} = U(g) s^* \sqrt{\sum_{k=1}^n w_{ks}^2}. \quad (19)$$

Здесь g - доверительная вероятность, $U(g)$ - квантиль нормального распределения порядка $\frac{1+g}{2}$. В формуле (19) $D^*(x^*(t))$ - состоятельная оценка дисперсии точечного прогноза $x^*(t)$. В соответствии с (18) она является произведением состоятельной оценки s^* среднего квадратического отклонения s случайных погрешностей E_k на известную статистику детерминированную функцию от t . Величину s^* рассчитывают согласно (10) и (15).

7. Пример применения непараметрического метода наименьших квадратов в модели с периодической составляющей

Обработаем фактические данные ОАО «Магнитогорский металлургический комбинат» о закупочных ценах на лом черных металлов [3]. Как

показано в [3], может быть использована модель (1) линейного тренда с периодической составляющей. Для облегчения расчетов оставим из каждого квартала данные только по одному месяцу. Введем условные моменты времени, а именно, будем измерять время в кварталах, начиная с первого квартала 2003 г. Исходные данные для демонстрации примера применения непараметрического метода наименьших квадратов в модели с периодической составляющей - пары чисел (t_k, x_k) , $k = 1, 2, \dots, 12$, - представлены в табл.1 в столбцах (3) и (4) соответственно.

По формулам (3) найдем оценки параметров a^* и d^* , что позволяет построить оценку трендовой составляющей

$$y^*(t) = a^*(t - \bar{t}) + d^* = 212,26 (t - 6,5) + 3967,17 = 212,26 t + 2587,48.$$

Численные значения трендовой составляющей приведены в столбце (5) табл.1.

Рассчитав отклонения исходных значений закупочных цен от оценок трендовой составляющей (столбец (6) табл.1), возведя их в квадрат и сложив, получаем остаточную сумму квадратов $SS = 4\,539\,214$ и $SS/n = SS/12 = 378\,267,843$.

Таблица 1

Построение модели прогнозирования цен на лом марки 3А

№ п/п	Периоды времени	Условные моменты времени	Закупочные цены, руб./т	Оценка тренда	Отклонения от оценки тренда	Восстановленные значения	Кажущиеся невязки
k		t_k	x_k	$y^*(t_k)$	$x_k - y^*(t_k)$	x_k^*	$x_k - x_k^*$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	янв.03	1	2 750	2 800	- 50	2 424	326
2	апр.03	2	3 800	3 012	788	3 545	255
3	июл.03	3	2 900	3 224	- 324	2 655	245

4	окт.03	4	3 100	3 437	- 337	3 848	- 748
5	январ.04	5	2 761	3 649	- 888	3 273	- 512
6	апр.04	6	4 602	3 861	741	4394	208
7	июл.04	7	3 540	4 073	- 533	3504	36
8	окт.04	8	5 268	4 286	982	4 697	571
9	январ.05	9	4 307	4 498	- 191	4 122	185
10	апр.05	10	4 779	4 710	69	5 243	- 464
11	июл.05	11	4 071	4 922	- 851	4 353	- 280
12	окт.05	12	5 723	5 135	588	5546	177

Сгруппировав отклонения исходных значений закупочных цен от оценок трендовой составляющей по месяцам (табл.2), наглядно убеждаемся в наличии периодической составляющей. Взяв среднее арифметическое отклонений от тренда за конкретный месяц, рассчитываем оценку $f^*(t_s)$ периодической составляющей (в соответствии с формулой (12)). Результаты приведены в табл.2.

Рассчитав по формуле (13) оценки периодической составляющей на весь интервал времени и сложив их с оценками трендовой составляющей, получаем в соответствии с формулой (14) оценку зависимости, «очищенную» от случайной составляющей, т.е. восстановленные значения (столбец (7) табл.1). Кажущиеся невязки, т.е. отклонения исходных значений закупочных цен от восстановленных значений, приведены в столбце (8) табл.1. Сравнивая столбцы (6) и (8), убеждаемся в целесообразности введения в модель периодической составляющей. В 9 случаях из 12 абсолютные величины отклонений уменьшились, в остальных трех, хотя и возросли, но лишь до среднего уровня среди остальных.

Таблица 2

Оценивание периодической составляющей

Номер квартала s	Месяц	Отклонения от тренда			Оценка $g_s^* = f^*(t_s)$ периодической составляющей
		В 2003 г.	В 2004 г.	В 2005 г.	
1	Январь	-50	- 888	-191	- 376
2	Апрель	788	741	69	533
3	Июль	- 324	- 533	- 851	- 569
4	Октябрь	- 337	982	588	411

Возведя в квадрат оценки периодической составляющей (табл.2), сложив эти квадраты, умножив на число лет и поделив на n , получаем, что $\frac{1}{n} \sum_{k=1}^n (f^*(t_k))^2 = 229\,537$. В соответствии с формулой (10) оценкой дисперсии случайной составляющей является

$$(s^*)^2 = \frac{SS}{n} - \frac{1}{n} \sum_{k=1}^n (f^*(t_k))^2 = 378\,267,83 - 229\,537 = 148\,731,$$

а оценкой среднего квадратического отклонения

$$s^* = \sqrt{148731} = 385,7.$$

В соответствии с формулами (4) и (5) оценим дисперсии оценок параметров

$$D^*(a^*) = \sum_{k=1}^n c_k^2 D^*(E_k) = \frac{(s^*)^2}{\sum_{k=1}^n (t_k - \bar{t})^2} = \frac{148731}{143} = 1040,$$

$$D^*(d^*) = \frac{(s^*)^2}{n} = \frac{148731}{12} = 12394.$$

Средние квадратические отклонения a^* и d^* оцениваются как 32,25 и 111,33 соответственно, а доверительные интервалы для доверительной вероятности 0,95 таковы: $[a_{\min}; a_{\max}] = [149,05; 275,47]$, $[d_{\min}; d_{\max}] = [3748,96; 4185,38]$.

Первое из условий (7) выполнено в силу построения оценок периодической составляющей по целому числу периодов. Действительно, согласно

данным табл.2 сумма оценок периодической составляющей для 12 точек наблюдений равна (-3), незначительное отклонение от 0 вызвано ошибками округления.

В соответствии с формулой (5) смещение оценки a^* оценивается как

$$\sum_{k=1}^n c_k f^*(t_k) = \frac{\sum_{k=1}^n (t_k - \bar{t}) f^*(t_k)}{\sum_{k=1}^n (t_k - \bar{t})^2} = \frac{5568}{143} = 38,94.$$

Таким образом, смещение имеет тот же порядок, что и среднее квадратичное отклонение оценки a^* , и заведомо меньше, чем полуширина доверительного интервала. Дальнейшее сравнение может быть проведено на основе оценки дисперсии смещения – случайной величины

$$Z = \frac{\sum_{k=1}^n (t_k - \bar{t}) f^*(t_k)}{\sum_{k=1}^n (t_k - \bar{t})^2}.$$

Алгоритм вычисления дисперсии Z аналогичен таковым для периодической составляющей и интервального прогноза (см. (16) и (18) соответственно), но более сложен, поэтому не включен в статью. Таким образом, можно считать, что предположения (7) модели (1) выполнены для данных табл.1.

Перейдем к оценке дисперсий значений периодической составляющей. Как следует из равенства (16),

$$D(g_s^*) = s^2 \sum_{k=1}^n h_{ks}^2, \quad s = 1, 2, \dots, q,$$

где $h_{ks} = -c_k r_s - \frac{1}{n} + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $h_{ks} = -c_k r_s - \frac{1}{n}$ при иных

значениях индекса суммирования k , и $r_s = \frac{1}{m} \sum_{j=1}^m (t_{s+(j-1)q} - \bar{t})$.

Начнем со значения $s = 1$ (периодическая составляющая для января).

Тогда $r_1 = \frac{1}{3}((1 - 6,5) + (5 - 6,5) + (9 - 6,5)) = -1,5$. Понадобятся значения

$$c_k = \frac{t_k - \bar{t}}{\sum_{k=1}^n (t_k - \bar{t})^2} = \frac{t_k - 6,5}{143} = \frac{k - 6,5}{143}.$$

Расчет удобно проводить с помощью таблицы (табл.3).

В табл. 3 столбец (3) получен из столбца (2) умножением на $\frac{r_1}{143} = \frac{-1,5}{143} = -0,01049$, каждый элемент столбца (6) равен сумма элементов столбцов (3), (4) и (5), стоящих в той же строке, а в столбце (7) стоят квадраты соседних элементов из столбца (6). Цель построения табл.3 – расчет суммы элементов столбца (7). Эта сумма равна 0,28275. Следовательно,

$$\sqrt{D^*(g_1^*)} = s * \sqrt{\sum_{k=1}^n h_{k1}^2} = 385,7 * \sqrt{0,28275} = 204,8.$$

Таблица 3

Расчет дисперсии периодической составляющей

<i>k</i>	$t_k - \bar{t}$	$c_k r_1$	$-1/n$	$+1/m$	h_{k1}	h_{k1}^2
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	- 5,5	0,0577	- 0,0833	0,3333	0,3077	0,09468
2	- 4,5	0,0472	- 0,0833	-	- 0,0361	0,00130
3	- 3,5	0,0367	- 0,0833	-	- 0,0466	0,00217
4	- 2,5	0,0262	- 0,0833	-	- 0,0571	0,00326
5	- 1,5	0,0157	- 0,0833	0,3333	0,2657	0,07060
6	- 0,5	0,0052	- 0,0833	-	- 0,0781	0,00610
7	0,5	- 0,0052	- 0,0833	-	- 0,0885	0,00783
8	1,5	- 0,0157	- 0,0833	-	- 0,0990	0,00980
9	2,5	- 0,0262	- 0,0833	0,3333	0,2238	0,05009
10	3,5	- 0,0367	-0,0833	-	0,1200	0,01440
11	4,5	- 0,0472	-0,0833	-	0,1305	0,01703
12	5,5	- 0,0577	-0,0833	-	0,1410	0,01988

Доверительный интервал для значения периодической составляющей в январе ($-376 - 1,96 \times 204,8$; $-376 + 1,96 \times 204,8$) захватывает 0 (при доверительной вероятности 0,95), отличие значения периодической составляющей от 0 не значимо (на уровне значимости 0,05).

Аналогичный расчет для значения $s = 2$ (периодическая составляющая для апреля) дает

$$\sum_{k=1}^n h_{k2}^2 = 0,25524, \quad \sqrt{D^*(g_2^*)} = s * \sqrt{\sum_{k=1}^n h_{k2}^2} = 385,7 \times \sqrt{0,25524} = 194,86.$$

Доверительный интервал для значения периодической составляющей в апреле ($533 - 1,96 \times 194,86$; $533 + 1,96 \times 194,86$) = ($533 - 381,93$; $533 + 381,93$) не захватывает 0 (при доверительной вероятности 0,95), отличие значения периодической составляющей от 0 значимо (на уровне значимости 0,05).

Приступим к завершающему этапу анализа данных табл.1 – построению интервального прогноза. Необходимо рассчитать величины $w_{ks} = c_k(t - \bar{t} - r_s) + \frac{1}{m}$, если $k \in \{s + (j-1)q, j = 1, 2, \dots, m\}$, и $w_{ks} = c_k(t - \bar{t} - r_s)$ при всех остальных значениях индекса суммирования k , где r_s – то же, что и в формуле (16), поскольку точечный прогноз $x^*(t)$ является несмещенным, асимптотически нормальным, а его дисперсия оценивается согласно (18) так:

$$D^*(x^*(t)) = (s^*)^2 \sum_{k=1}^n w_{ks}^2.$$

Начнем с прогноза на январь 2006 г. (по данным за 2003 - 2005 гг.). Тогда $t = 13$, $s = 1$, $r_1 = -1,5$, $w_{k1} = 8c_k + \frac{1}{3}$, если $k \in \{1 + 4(j-1), j = 1, 2, 3\}$, и $w_{k1} = 8c_k$ при всех остальных значениях индекса суммирования. При этом $8c_k = 8 \frac{k - 6,5}{143} = \frac{8k - 52}{143}$. Расчет удобно проводить с помощью таблицы (табл.4).

Сумма значений, стоящих в последнем столбце табл.4, равна 0,61299.
Согласно формуле (19)

$$\Delta(13) = U(0,95)\sqrt{D^*(x^*(13))} = 1,96 \times 385,7 \times \sqrt{0,61299} = 591,88.$$

Согласно (14) точечный прогноз прогностической функции таков:

$$x^*(13) = a^*(13 - \bar{t}) + d^* + f^*(13) = 212,26 \times 13 + 2587,48 + (-376) = 4971.$$

Нижняя и верхняя доверительные границы для прогностической функции (с учетом как трендовой, так и периодической составляющих) имеют вид:

$$x_{нижн}(13) = 4971 - 592 = 4379, \quad x_{верх}(13) = 4971 + 592 = 5563.$$

Таблица 4

Расчет дисперсии прогностической функции

k	$\frac{8k - 52}{143}$	$1/m$	w_{k1}	w_{k1}^2
1	- 0,3077	0,3333	0,0256	0,00066
2	- 0,2517	-	- 0,2517	0,06336
3	- 0,1958	-	- 0,1958	0,03834
4	- 0,1399	-	- 0,1399	0,01957
5	- 0,0839	0,3333	0,2494	0,06220
6	- 0,0280	-	- 0,0280	0,00078
7	0,0280	-	0,0280	0,00078
8	0,0839	-	0,0839	0,00700
9	0,1399	0,3333	0,4732	0,22392
10	0,1958	-	0,1958	0,03834
11	0,2517	-	0,2517	0,06336
12	0,3077	-	0,3077	0,09468

Реальное значение (табл.4.1.7) - 4336. Оно практически совпадает с нижней доверительной границей прогностической функции $x_{нижн}(13)$.

Аналогичные расчеты для апреля 2006 г. ($t = 14, s = 2, r_2 = -0,5$) дают $\Delta(14) = 1,96 \times 385,7 \times \sqrt{0,72480} = 643,60$. Точечный прогноз равен $x^*(14) = 6092$, а нижняя и верхняя доверительные границы таковы: $x_{нижн}(14) = 5448, x_{верх}(14) = 6736$. Реальное значение (табл.4.1.7) – 5430. Оно практически совпадает с нижней доверительной границей прогностической функции $x_{нижн}(14)$.

8. Интервальный прогноз индивидуальных значений

Формула (19) позволяет строить интервальный прогноз для прогностической функции, т.е. для математического ожидания временного ряда. Наблюдаемое значение отличается от него на величину невязки. Распределение невязки можно оценить по значениям кажущихся невязок (см. столбец (8) в табл.1). Напомним, что это распределение не является нормальным, не описывается элементом какого-либо параметрического семейства. Интервальный прогноз индивидуального значения построить, скорректировав интервальный прогноз для прогностической функции с помощью выборочных квантилей кажущихся невязок.

Для рассмотренного выше примера вариационный ряд $n = 12$ кажущихся невязок таков: -748, - 512, - 464, - 280, 36, 177, 185, 208, 245, 255, 326, 571. Нижний дециль оценим как второй член вариационного ряда (- 512), верхний – как предпоследний (одиннадцатый) член вариационного ряда 326. Для расчета нижней доверительной границы индивидуального значения надо взять нижнюю доверительную границу прогностической функции и отнять 512. Для расчета верхней доверительной границы индивидуального значения надо взять верхнюю доверительную границу прогностической функции и прибавить 326.

Итак, для данных табл.1 индивидуальные значения лежат «глубоко внутри» доверительных интервалов. Прогнозы полностью оправдались.

9. О проверке условий (7)

Рассмотрим три вопроса. Верны ли условия (7) в моделях, соответствующих реальным ситуациям? Как проверять справедливость условий по результатам наблюдений? Каковы свойства оценок, если эти условия оказываются невыполненными?

В условиях (7) важную роль играет система точек наблюдения t_k , $k = 1, 2, \dots, n$. Более тщательно рассмотрим ранее принятую *модель с целым числом периодов*, для которой справедливо соотношение (11). При этом объем наблюдений $n = tq$, где t – количество периодов, q – число наблюдений в одном периоде. Предполагается, что первые q моментов наблюдения при сдвиге на длину периода дают следующие q моментов времени, при сдвиге на две длины периода дают третий набор из q моментов наблюдения, и т.д. Для значений периодической составляющей выше построены точечные оценки и доверительные интервалы (в предположении, что количество периодов t безгранично растет), в чем и состоит оценивание периодической составляющей. (Для гладкой функции $f(t)$ при безграничном росте числа наблюдений q в одном периоде можно получить сходимость оценок периодической составляющей не только в q точках, но и на всем периоде. При этом от оценок в q точках придется перейти к оценкам на всем периоде, например, кусочно-линейным, соединив соседние точки графика отрезками прямых.)

Описанная модель справедлива, когда, например, в течение некоторого числа лет имеются поквартальные или ежемесячные данные бухгалтерского учета. При изучения посещений сайта или торгового заведения – почасовые данные за целое число недель. Если в ряду наблюдений есть пропуски (временной ряд не является полным) – предпосылки модели не вы-

полняются. Если система точек наблюдения не образует арифметическую прогрессию,

В рассматриваемой модели естественно принять, что

$$\sum_{k=1}^q f(t_k) = 0, \quad (20)$$

суммарное отклонение значений восстанавливаемой функции от линейного тренда за один период является нулевым. Тогда первое из условий (7) выполнено:

$$\sum_{i=1}^n f(t_i) = m \sum_{k=1}^q f(t_k) = 0 .$$

В реальных ситуациях система точек наблюдения может включать в себя, кроме целого числа периодов, еще несколько начальных точек следующего периода. Можно априори принять первое условие (7), для этого изменив – при необходимости – величину свободного члена d в модели тренда (та же логика рассуждений, что и при принятии условий $M(e_k) = 0$ – в модели без периодической составляющей – и $M(E_k) = 0$ в общем случае). Однако возникает противоречие между первым условием (7) и условием (20). Условие первое условие (7) автоматически обеспечивается методом наименьших квадратов, а условие (20) соответствует логике моделирования. Однако поскольку рассматриваем асимптотическую теорию при безграничном росте числа периодов, указанное различие исчезает при $m \rightarrow \infty$. Таким образом, первое из условий (7) вытекает из свойств рассматриваемой модели и потому вообще не требует проверки по экспериментальным данным, в отличие от второго условия (7), которое выполнено не всегда.

Добавим к модели с целым числом периодов два предположения - симметричности множества $\{t_k, k = 1, 2, \dots, n\}$ относительно \bar{t} и четности периодической составляющей $f(t)$ относительно той же точки. Эти предположения выполнены, если, например, график $f(t)$ симметричен относительно середины года. Тогда второе условие (7) выполнено. Ясно, что обычно

нет оснований априори считать, что реальные данные описываются такой моделью.

10. Проверка второго условия (7) по экспериментальным данным

Естественно использовать статистику

$$Y = \sum_{j=1}^n (t_j - \bar{t}) f^*(t_j),$$

где $f^*(t_j)$ - ранее построенная оценка периодической составляющей $f(t)$. Оценка $f^*(t_j)$ является несмещенной, а потому

$$M(Y) = \sum_{j=1}^n (t_j - \bar{t}) f(t_j).$$

При справедливости (6) распределение Y является асимптотически нормальным (при безграничном росте количества периодов m). Для проверки второго условия (7), т.е. для проверки нулевой гипотезы $H_0: M(Y) = 0$ при альтернативной гипотезе о неравенстве математического ожидания 0 достаточно оценить дисперсию Y .

В соответствии с (11) формулу (16) можно записать для любого $j = 1, 2, \dots, n$, если под $k = k(j)$ понимать $k(j) = j - aq$ при максимально возможном a , при котором $k(j)$ остается положительным, т.е. $k(j)$ – это остаток от деления j на q , если этот остаток ненулевой, и $k(j) = q$ при нулевом остатке. Таким образом,

$$f^*(t_j) = f(t_j) + \sum_{i=1}^n h_{ik(j)} E_i, \quad j = 1, 2, \dots, n, \quad (21)$$

где h_{ik} – те же, что и в формуле (16). В соответствии с определением Y из (21) следует, что

$$Y = \sum_{j=1}^n (t_j - \bar{t}) f^*(t_j) = \sum_{j=1}^n (t_j - \bar{t}) f(t_j) + \sum_{j=1}^n \left[(t_j - \bar{t}) \sum_{i=1}^n h_{ik(j)} E_i \right]. \quad (22)$$

Изменим порядок суммирования во втором слагаемом в (22):

$$\sum_{j=1}^n \left[(t_j - \bar{t}) \sum_{i=1}^n h_{ik(j)} E_i \right] = \sum_{i=1}^n \left[\sum_{j=1}^n (t_j - \bar{t}) h_{ik(j)} \right] E_i .$$

Следовательно, поскольку E_i – независимые одинаково распределенные случайные величины с математическим ожиданием 0 и дисперсией s^2 , то

$$D(Y) = \sum_{i=1}^n \left[\sum_{j=1}^n (t_j - \bar{t}) h_{ik(j)} \right]^2 s^2 . \quad (23)$$

Величину s^2 оцениваем по формулам (10) и (15), величины $h_{ik(j)}$ описаны после формулы (16). Подставив оценку s^2 в (23), получаем оценку $D^*(Y)$ дисперсии Y .

В соответствии с асимптотической нормальностью Y правило принятия решений при проверке гипотезы $H_0: M(Y) = 0$ таково: если

$$\left| \frac{Y}{\sqrt{D^*(Y)}} \right| \leq C(a) = \Phi^{-1} \left(1 - \frac{a}{2} \right), \quad (24),$$

где $C(a)$ - критическое значение, соответствующее уровню значимости a , то нулевая гипотеза принимается (второе условие (7) выполнено), если же неравенство (24) не выполнено, то принимается альтернативная гипотеза (второе условие (7) не выполнено).

11. Асимптотическая несмещенность оценки параметра a

Приведем пример, когда второе условие (7) не выполнено. Измерять время будем в месяцах. Пусть данные берутся на середину квартала. Тогда последовательность моментов времени такова: 2, 5, 8, 11, 14, 17, 20, 23, ... Задан период – год. Периодическая составляющая задается четырьмя числами: $g_1 = -1$, $g_2 = -2$, $g_3 = -3$, $g_4 = 6$. Для таких данных выполнено равенство (20), т.е. $\sum_{k=1}^q f(t_k) = -1 - 2 - 3 + 6 = 0$. Следовательно, выполнено первое условие (7). Используя это условие, можно упростить второе условие (7):

$$\sum_{i=1}^n (t_i - \bar{t}) f(t_i) = \sum_{i=1}^n t_i f(t_i) - \bar{t} \sum_{i=1}^n f(t_i) = \sum_{i=1}^n t_i f(t_i) = 0.$$

Для простоты расчетов ограничимся двумя годами. Тогда

$$\begin{aligned} \sum_{i=1}^n t_i f(t_i) &= 2 \times (-1) + 5 \times (-2) + 8 \times (-3) + 11 \times 6 + 14 \times (-1) + 17 \times (-2) + 20 \times (-3) + \\ &+ 23 \times 6 = (-2) + (-10) + (-24) + 66 + (-14) + (-34) + (-60) + 138 = 60 \end{aligned}$$

Второе условие (7) не выполнено. Оно не будет выполнено и для любого иного числа лет. Действительно, если x – начало года (для первого года $x = 0$, для второго $x = 12$, и т.д.), то вклад этого года в рассматриваемую сумму будет равен

$$\begin{aligned} (x+2) \times (-1) + (x+5) \times (-2) + (x+8) \times (-3) + (x+11) \times 6 = \\ = 2 \times (-1) + 5 \times (-2) + 8 \times (-3) + 11 \times 6 = 30 \end{aligned}$$

Причина нарушения второго условия (7) ясна – периодическая составляющая не симметрична в течение года. Такое поведение периодической составляющей естественно для сельскохозяйственных предприятий. Противоположную ситуацию демонстрирует периодическая составляющая для временного ряда цен на лом черных металлов (по данным Магнитогорского металлургического комбината), проанализированного выше.

Смещение оценки параметра a равно

$$M(a^*) - a = \frac{\sum_{i=1}^n f(t_i)(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2} = \frac{\sum_{i=1}^n f(t_i)t_i}{\sum_{i=1}^n (t_i - \bar{t})^2}. \quad (25)$$

В рассматриваемом примере числитель за m лет равен $30m$. А знаменатель, очевидно, имеет порядок m^3 . Смещение имеет порядок m^{-2} , т.е. быстро убывает с ростом числа периодов. Оценка a^* параметра a является асимптотически несмещенной.

Нетрудно показать, что для модели с целым числом периодов всегда имеет асимптотическая несмещенность оценки a^* параметра a . Если второе условие (7) выполнено – эта оценка является несмещенной, если не выполнено – смещенной, но смещение стремится к 0 при росте числа пе-

риодов. Таким образом, выполнение второго условия (7) не является необходимым для применения рассматриваемых методов. Тем не менее проверка второго условия (7) по экспериментальным данным является полезным для решения о том, можно ли пользоваться асимптотической несмещенностью оценки при имеющемся объеме данных.

12. Заключительные замечания

Подведем итоги. По сравнению с эвристическими алгоритмами, разобранными в [1] и других литературных источниках, разработанная в настоящей статье теория позволила:

1) дать общее обоснование этим алгоритмам в рамках асимптотических методов математической статистики и указать условия их применимости (формула (6));

2) выявить принципиально важные условия (7), необходимые и достаточные для несмещенности и состоятельности рассматриваемых оценок;

3) построить доверительные интервалы для зависимости (прогностической функции), трендовой и периодической составляющих, индивидуальных значений временного ряда.

Обсуждение отдельных сторон рассматриваемой проблемы проведено в работах [2, 4, 5].

В рамках математической статистики удастся провести анализ не всех распространенных эвристических алгоритмов. Так, довольно часто рекомендуют вначале провести сглаживание («выравнивание») временного ряда, например, методом скользящих средних [1, с.137]. При этом периодическая (сезонная) составляющая меняется (также сглаживается), а погрешности (отклонения от суммы трендовой и периодической составляющих) становятся зависимыми случайными величинами, что делает невозможным применение описанных в настоящей статье методов.

Теория устойчивости [6] отвергает идею поиска оптимального метода, поскольку зачастую оказывается, что для любого выбранного для рассмотрения метода анализа данных можно подобрать такое понимание оптимальности, что именно этот метод является оптимальным. Например, метод наименьших квадратов в определенном смысле оптимален, если погрешности имеют нормальное распределение, в то время как метод наименьших модулей оптимален, если погрешности имеют распределение Лапласа. В задаче проверки однородности двух независимых выборок установлено [7], что для любого из обычно используемых критериев однородности существует такое распределение на множестве альтернативных гипотез, что рассматриваемый критерий является оптимальным (в том смысле, который определен в [7]).

Работа выполнена в рамках новой парадигмы прикладной (математической) статистики [8, 9]. Изучена *непараметрическую модель*, не предполагающая, что распределение погрешностей (ошибок, невязок) входит в то или иное параметрическое семейство. Второе основание для того, чтобы именовать рассматриваемые модель и метод непараметрическими – оценивание периодической составляющей произвольного вида, т.е. отказ от выбора периодической составляющей из какого-либо параметрического семейства функций.

Полученные в статье [10] научные результаты, касающиеся средних величин и законов больших чисел в пространствах произвольной природы, могут быть применены для анализа данных в различных научных и прикладных областях. В отличие от них результаты настоящей работы нацелены прежде всего на анализ динамических рядов экономических показателей (временных рядов), необходимость которого часто возникает при организационно-экономическом моделировании с целью решения задач управления хозяйственными единицами [11]. Именно потребности экономики и управления ставят во главу угла модели с одной независимой пе-

ременной – временем. Длина периода задается существом рассматриваемой прикладной задачи (для оценки длины периода по статистическим данным нужен другой математический аппарат, разработанный в [12]). Рассмотренные в настоящей работе постановки можно относить к эконометрике [2], т.е. статистическим методам в экономике. Полученные результаты могут быть применены для прогнозирования и построения экономико-математических моделей, в частности, в рамках неформальной информационной экономики будущего [13].

Литература

1. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др.; Под ред. И.И. Елисеевой. – М.: Финансы и статистика. 2001. – 192 с.
2. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. – Ростов-на-Дону: Феникс, 2009. - 572 с.
3. Крюкова Е.М. Применение методов организационно-экономического прогнозирования в отрасли лома черных металлов // Заводская лаборатория. Диагностика материалов. - 2008. - Т.74. - №7. - С.67 – 72.
4. Орлов А.И. Непараметрический метод наименьших квадратов: учет сезонности // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып. XXI. – Пермь: Перм. ун-т, 2008. – С.135-148.
5. Орлов А.И. Непараметрический метод наименьших квадратов с периодической составляющей: условия применимости // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып. XXII. – Пермь: Перм. ун-т, 2010. – С.96-108.
6. Орлов А.И. Устойчивые экономико-математические методы и модели. Разработка и развитие устойчивых экономико-математических методов и моделей для модернизации управления предприятиями. – Saarbrücken: LAP, 2011. – 436 с.
7. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
8. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. - 2012. - Том 78. - №1, часть I. - С.87-93.
9. Орлов А.И. Основные черты новой парадигмы математической статистики / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С.188-214. – IDA [article ID]: 0901306013. – Режим доступа: <http://ej.kubagro.ru/2013/06/pdf/13.pdf>, 1,688 у.п.л.
10. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(89). – С. 554 – 584. IDA [article ID]: 0891304038. – Режим доступа: <http://ej.kubagro.ru/2013/05/pdf/38.pdf>, 1,938 у.п.л., импакт-фактор РИНЦ=0,577

11. Орлов А.И. Организационно-экономическое моделирование при решении задач управления хозяйственными единицами / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №03(87). – С. 679–705. – Режим доступа: <http://ej.kubagro.ru/2013/03/pdf/53.pdf>

12. Орлов А.И. Метод оценивания длины периода и периодической составляющей сигнала // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. – Пермь: Изд-во Пермского государственного университета, 1999. – С.38-49.

13. Орлов А.И. Проблемы методологии государственной политики и управления в неформальной информационной экономике будущего / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(88). – С. 592 – 618. – IDA [article ID]: 0881304041. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/41.pdf>

References

1. Praktikum po jekonometrike: Ucheb. posobie / I.I. Eliseeva, S.V. Kurysheva, N.M. Gordeenko i dr.; Pod red. I.I. Eliseevoj. – М.: Finansy i statistika. 2001. – 192 s.

2. Orlov A.I. Jekonometrika. Izd. 4-e, dop. i pererab. Uchebnik dlja vuzov. – Rostov-na-Donu: Feniks, 2009. - 572 s.

3. Krjukova E.M. Primenenie metodov organizacionno-jekonomicheskogo prognozirovanija v otrasli loma chernyh metallov // Zavodskaja laboratorija. Diagnostika materialov. - 2008. - T.74. - №7. - S.67 – 72.

4. Orlov A.I. Neparаметричeskij metod naimen'shij kvadratov: uchet se-zonnosti // Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. Vyp. XXI. – Perm': Perm. un-t, 2008. – S.135-148.

5. Orlov A.I. Neparаметричeskij metod naimen'shij kvadratov s periodicheskoj sostavljajushhej: uslovija primenimosti // Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. Vyp. XXII. – Perm': Perm. un-t, 2010. – S.96-108.

6. Orlov A.I. Ustojchivye jekonomiko-matematicheskie metody i modeli. Razrabotka i razvitie ustojchivykh jekonomiko-matematicheskikh metodov i modelej dlja modernizacii upravlenija predpriyatijami. – Saarbrucken: LAP, 2011. – 436 s.

7. Nikitin Ja.Ju. Asimptoticheskaja jeffektivnost' neparаметричeskikh kriteriev. - М.: Nauka, 1995. - 240 s.

8. Orlov A.I. Novaja paradigma prikladnoj statistiki // Zavodskaja laboratorija. Diagnostika materialov. - 2012. - Tom 78. - №1, chast' I. - S.87-93.

9. Orlov A.I. Osnovnye cherty novoj paradigmy matematicheskoi statistiki / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Краснодар: KubGAU, 2013. – №06(090). S.188-214. – IDA [article ID]: 0901306013. – Rezhim dostupa: <http://ej.kubagro.ru/2013/06/pdf/13.pdf>, 1,688 u.p.l.

10. Orlov A.I. Srednie velichiny i zakony bol'shij chisel v prostranstvah proizvol'noj prirody / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Краснодар: KubGAU, 2013. – №05(89). – S. 554 – 584. IDA [article ID]: 0891304038. – Rezhim dostupa: <http://ej.kubagro.ru/2013/05/pdf/38.pdf>, 1,938 u.p.l., impakt-faktor RINC=0,577

11. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie pri reshenii zadach upravlenija hozjajstvennymi edinicami / A.I. Orlov // Politematicheskij setevoj jel-

elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №03(87). – S. 679–705. – Rezhim dostupa: <http://ej.kubagro.ru/2013/03/pdf/53.pdf>

12. Orlov A.I. Metod ocenivaniya dliny perioda i periodicheskoy sostav-ljajushhej signala // Statisticheskie metody ocenivaniya i proverki gipotez. Mezhvuzovskij sbornik nauchnyh trudov. – Perm': Izd-vo Permskogo gosudarstvennogo universiteta, 1999. – S.38-49.

13. Orlov A.I. Problemy metodologii gosudarstvennoj politiki i upravlenija v neformal'noj informacionnoj jekonomike budushhego / A.I. Orlov // Polite-maticheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №04(88). – S. 592 – 618. – IDA [article ID]: 0881304041. – Rezhim dostupa: <http://ej.kubagro.ru/2013/04/pdf/41.pdf>