Математические методы исследования

УДК 519.2

НЕПАРАМЕТРИЧЕСКИЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ С ПЕРИОДИЧЕСКОЙ СОСТАВЛЯЮЩЕЙ (обобщающая статья)

© А. И. Орлов[1]

Статья поступила 15 января 2013 г.

Рассмотрена непараметрическая задача восстановления зависимости, которая описывается суммой линейного тренда и периодической функции с известным периодом. Получены асимптотические распределения оценок параметров и трендовой составляющей. Найдено математическое ожидание остаточной суммы квадратов. 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.

Ключевые слова: метод наименьших квадратов; непараметрические методы; периодическая составляющая; оценивание; прогнозирование.

Метод наименьших квадратов восстановления зависимости — один из наиболее распространенных математических методов исследования. В данной работе рассмотрена непараметрическая постановка: восстанавливаемая зависимость — сумма линейной функции и периодической составляющей произвольного вида (с известным периодом), с произвольным распределением случайных погрешностей (остатков, невязок).

Задача восстановления линейной зависимости

Начнем с простейшего случая — 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), $k = 1, 2, ..., n$, 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, например, в работе [1], аддитивную модель $T + S + E$ и мультипликативную модель $T \times S \times E$.

Простейшая аддитивная модель имеет вид

$$x_k = a(t_k - \bar{t}) + d + e_k = a(t_k - \bar{t}) + d + f(t_k) + E_k,$$

$$k = 1, 2, ..., n. \qquad (1)$$

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 σ^2, неизвестной статистику. В рассматриваемой модели $e_k = f(t_k) + E_k$, $k = 1, 2, ..., n$, 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

$$y_k = [Bt_k^a] f_1(t_k)(1 + \varepsilon_k),\ k = 1, 2, ..., n. \qquad (2)$$

где смысл сомножителей описан выше. При логарифмировании модель (2) переходит в аналог модели (1), следовательно, достаточно рассматривать модель (1).

Иногда принимают предположение о нормальности распределения погрешностей. Однако известно,

[1] Институт высоких статистических технологий и эконометрики Московского государственного технического университета им. Н. Э. Баумана; Московский физико-технический институт, Москва, Россия; e-mail: prof-orlov@mail.ru

что распределения реальных данных, как правило, отличаются от нормальных [2]. Поэтому далее будем рассматривать 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 данные 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 работе [1], являются эвристическими. Цель данной работы — *построение непараметрической вероятностно-статистической теории прогноза временного ряда на базе линейного тренда с учетом аддитивной периодической составляющей.*

Метод наименьших квадратов разработан К. Гауссом в 1794 г. [2]. Согласно этому методу для получения наилучшей функции, выражающей 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 (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

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

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

Асимптотические распределения оценок параметров

Из формулы (3) следует, что

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

$$=d+\frac{1}{n}\sum_{k=1}^{n}e_k=d+\frac{1}{n}\sum_{k=1}^{n}f(t_k)+\frac{1}{n}\sum_{k=1}^{n}E_k. \qquad (4)$$

Согласно Центральной предельной теореме (для выполнения ее условий необходимо предположить, например, что погрешности e_k, $k = 1, 2, ..., n$, 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 σ^2/n. Из формул (3) и (4) следует, 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 нуль, поэтому

$$a^*=a+\sum_{k=1}^{n}c_k e_k=a+\sum_{k=1}^{n}c_k f(t_k)+\sum_{k=1}^{n}c_k E_k,$$

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

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

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

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

$$\frac{\lim_{n\to\infty}\max|t_k-\bar{t}|}{\sqrt{\sum_{k=1}^{n}(t_k-\bar{t})^2}}=0. \qquad (6)$$

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

Итак, дисперсии оценок МНК параметров a^* и d^* линейного тренда те же, что и при отсутствии сезонных искажений (см., например, [2]), а их математи-

ческие ожидания зависят от периодической составляющей. Однако в случае

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

оценки a^* и d^* являются несмещенными.

Условия (7) необходимы и достаточны для несмещенности и состоятельности оценок МНК коэффициентов линейной зависимости. Проверка условий (7) будет рассмотрена в конце работы.

Несмещенность (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 определенным значениям, прежде всего нулю.

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 $y^*(t) = a^*(t_k - \bar{t}) + d^*$ трендовой составляющей $y(t) = a(t - \bar{t}) + d$ рассматриваемой зависимости является несмещенной. 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

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

$$= \frac{1}{n} (t - \bar{t}) \sigma^2 \sum_{k=1}^{n} c_k = 0.$$

Таким образом,

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

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

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

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

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

где полуширина доверительного интервала

$$\delta(t) = U(\gamma) \sqrt{D^*[y^*(t)]} = U(\gamma) \sigma^* \sqrt{\frac{1}{n} + \frac{(t - \bar{t})^2}{\sum_{k=1}^{n} (t_k - \bar{t})^2}}. \qquad (9)$$

Здесь γ — доверительная вероятность; $U(\gamma)$ — квантиль нормального распределения порядка $(1 + \gamma)/2$, т.е. $U(\gamma) = \Phi^{-1}[(1 + \gamma)/2]$, где $\Phi(x)$ — функция стандартного нормального распределения с математическим ожиданием нуль и дисперсией единица. При $\gamma = 0{,}95$ (наиболее применяемое значение) имеем $U(\gamma) = 1{,}96$. В формуле (9) $D^*[y^*(t)]$ — состоятельная оценка дисперсии $y^*(t)$, которая в соответствии с выражением (8) является произведением состоятельной оценки σ^* среднего квадратического отклонения σ случайных погрешностей E_k на известную исследователю детерминированную функцию от t.

Математическое ожидание остаточной суммы квадратов

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

$$SS = \sum_{k=1}^{n} [y^*(t_k) - x_k]^2 =$$

$$= \sum_{k=1}^{n} \left[(a^* - a)(t_k - \bar{t}) + (d^* - d) - f(t_k) - E_k \right]^2.$$

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

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

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

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

Найдем математическое ожидание каждого из слагаемых:

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

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

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

$$\times[f(t_k) + E_k] + M[f(t_k) - E_k]^2.$$

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

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

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

$$= -2\left[c_k(t_k - \bar{t}) + \frac{1}{n}\right]\sigma^2,$$

$$M[f(t_k) - E_k]^2 = f^2(t_k) + \sigma^2.$$

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

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

Следовательно,

$$M\left(\frac{SS}{n}\right) = \sigma^2 + \frac{1}{n}\sum_{k=1}^{n} f^2(t_k), \qquad (10)$$

где первое слагаемое в правой части соответствует вкладу случайной составляющей, второе — вкладу периодической составляющей; в некоторых случаях второе слагаемое может быть известно из предыдущего опыта или же оценено экспертами.

Оценивание периодической составляющей

В литературе рассматривают как параметрические, так и непараметрические подходы. Согласно популярному методу достаточно гладкую функцию можно разложить в ряд Фурье и получить хорошее приближение с помощью небольшого числа гармоник, в простейшем случае — одной гармоники. Так, динамику индекса инфляции можно попытаться изучать с помощью модели

$$x_k = a(t_k - \bar{t}) + d + f(t_k) + E_k =$$

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

где время t измеряется в годах; неизвестные параметры a, 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 соответствии с предположениями (7) рассматривается целое число периодов, т.е. $n = mq$, где n — объем наблюдений, m — 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 — третий набор из q моментов наблюдения, и т.д. Тогда в соответствии с определением периодической составляющей справедливы равенства

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

Если наблюдения проводятся ежемесячно в течение m лет, то число наблюдений в одном периоде $q = 12$, общий объем наблюдений $n = 12m$, $s = 1, 2, ..., 12$.

Пусть g_s — общее значение в (11), тогда для оценки периодической составляющей требуется оценить $g_1, g_2, ..., g_q$. Естественный подход состоит в том, чтобы усреднить m значений $\bar{x}_k - \bar{y}^*(t_k)$, соответствующих моментам времени, отстоящим друг от друга на целое число периодов. Другими словами, следует усреднить «очищенные» от трендовой составляющей исходные данные, соответствующие одноименным месяцам различных лет. Речь идет об оценках

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

Оценка периодической составляющей распространяется на весь интервал наблюдений очевидным образом:

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

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

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

где 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 $\{t_k,\ k = 1, 2, ..., n\}$ 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 условиях суммы (12) периодическую составляющую оценивают состоятельно (при $m \to \infty$) и несмещенно. Как следствие,

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

по вероятности при $n \to \infty$. В соответствии с формулой (10) последнее соотношение дает возможность оценить σ^2, а затем построить интервальный прогноз для трендовой составляющей согласно выражению (9). В рассматриваемой ситуации, как правило, n растет, увеличиваясь на величины, кратные q — числу наблюдений в одном периоде. В результате уменьшаемое в (15) — константа, т.е. не зависит от n. Эти особенности связаны с тем, что выполнение условий (7) предполагает рассмотрение целого числа периодов.

Рассмотрим оценки (12) подробнее. Как следует из формул (4), (5), (11) и (12)

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

$$+ \frac{1}{m}\sum_{j=1}^{m} E_{s+(j-1)q},\ s = 1, 2, ..., q.$$

С учетом выражений (4), (5) и (7) получаем

$$g_s^* = f(t_s) - \left(\sum_{k=1}^{n} c_k E_k\right)\left[\frac{1}{m}\sum_{j=1}^{m}(t_{s+(j-1)q} - \bar{t})\right] -$$

$$- \frac{1}{n}\sum_{k=1}^{n} E_k + \frac{1}{m}\sum_{j=1}^{m} E_{s+(j-1)q},\ s = 1, 2, ..., q.$$

Таким образом,

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

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

$$r_s = \frac{1}{m}\sum_{j=1}^{m}(t_{s+(j-1)q} - \bar{t}).$$

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)].

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

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

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

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

где $w_{ks} = c_k(t - \bar{t} - r_s) + 1/m$, если $k \in \{s + (j-1)q, j = 1, 2, ..., m\}$, и $w_{ks} = c_k(t - \bar{t} - r_s)$ при всех остальных значениях индекса суммирования k, 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

$$D[x(t)] = \sum_{k=1}^{n} w_{ks}^2 D(E_k) = \sigma^2 \sum_{k=1}^{n} w_{ks}^2. \qquad (18)$$

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 следующий вид:

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

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

где

$$\Delta(t) = U(\gamma)\sqrt{D^*[x^*(t)]} = U(\gamma)\sigma^*\sqrt{\sum_{k=1}^{n} w_{ks}^2}. \qquad (19)$$

Здесь γ — доверительная вероятность; $U(\gamma)$ — квантиль нормального распределения порядка $(1 + \gamma)/2$; $D^*[x^*(t)]$ — состоятельная оценка дисперсии точечного прогноза $x^*(t)$. Последняя в соответствии с выражением (18) является произведением состоятельной оценки σ* среднего квадратического отклонения σ случайных погрешностей E_k на известную статистику детерминированную функцию от t. Величину σ* рассчитывают согласно формулам (10) и (15).

Приведем *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]. Для этого 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), $k = 1, 2, ..., 12$ (табл. 1).

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

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

$$= 212{,}26t + 2587{,}48.$$

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

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

Таблица 1. Данные для построения модели прогнозирования цен на лом марки 3А

№ п/п k	Период	Условные моменты времени t_k	Закупочные цены, руб./т x_k	Оценка тренда $y^*(t_k)$	Отклонения от оценки тренда $x_k - y^*(t_k)$	Восстановленные значения x_k^*	Кажущиеся невязки $x_k - x_k^*$
1	Янв. 03	1	2750	2800	–50	2424	326
2	Апр. 03	2	3800	3012	788	3545	255
3	Июл. 03	3	2900	3224	–324	2655	245
4	Окт. 03	4	3100	3437	–337	3848	–748
5	Янв. 04	5	2761	3649	–888	3273	–512
6	Апр. 04	6	4602	3861	741	4394	208
7	Июл. 04	7	3540	4073	–533	3504	36
8	Окт. 04	8	5268	4286	982	4697	571
9	Янв. 05	9	4307	4498	–191	4122	185
10	Апр. 05	10	4779	4710	69	5243	–464
11	Июл. 05	11	4071	4922	–851	4353	–280
12	Окт. 05	12	5723	5135	588	5546	177

Таблица 2. Данные для оценивания периодической составляющей

Номер квартала s	Месяц	Отклонения от тренда			Оценка $g_s^* = f^*(t_s)$ периодической составляющей
		2003 г.	2004 г.	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

«очищенную» от случайной составляющей, т.е. восстановленные значения (см. табл. 1). Кажущиеся невязки, т.е. отклонения исходных закупочных цен от восстановленных, также приведены в табл. 1. Сравнивая отклонения от оценки тренда и кажущиеся невязки, убеждаемся в целесообразности введения в модель периодической составляющей. В девяти случаях из 12 абсолютные величины отклонений уменьшились, в остальных трех хотя и возросли, но лишь до среднего уровня среди остальных.

Возведя в квадрат оценки периодической составляющей (см. 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

$$(\sigma^*)^2 = \frac{SS}{n} - \frac{1}{n}\sum_{k=1}^{n}[f^*(t_k)]^2 =$$

$$= 378\,267{,}83 - 229\,537 = 148\,731,$$

а оценкой среднего квадратического отклонения

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

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

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

$$D^*(d^*) = (\sigma^*)^2/n = 148\,731/12 = 12\,394.$$

Средние квадратические отклонения a^* и d^* оцениваются как 32,25 и 111,33, а доверительные интервалы для доверительной вероятности 0,95 составляют:

$$[a_{\min};\ a_{\max}] = [149{,}05;\ 275{,}47],$$

$$[d_{\min};\ d_{\max}] = [3748{,}96;\ 4185{,}38].$$

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 нуля вызвано ошибками округления.

В соответствии с формулой (5) смещение оценки a^* находится 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 (16) и интервального прогноза (18), но более сложен, поэтому не включен в статью. Таким образом, можно считать, что предположения (7) модели (1) выполнены для данных табл. 1.

Таблица 3. Данные для расчета дисперсии периодической составляющей

k	$t_k - \bar{t}$	$c_k r_1$	$-1/n$	$+1/m$	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

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

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

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

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

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

Схему расчета удобно показать с помощью табл. 3. Здесь столбец *3* получен умножением столбца *2* на $r_1/143 = -1{,}5/143 = -0{,}01049$, каждый элемент столбца *6* равен сумме элементов столбцов *3*, *4* и *5*, стоящих в той же строке, а в столбце *7* указаны квадраты соседних элементов из столбца *6*. Сумма элементов столбца *7* равна 0,28275. Следовательно,

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

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

Таблица 4. Данные для расчета дисперсии прогностической функции

k	$8k - 52/143$	$1/m$	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

Аналогичный расчет для $s = 2$ (периодическая составляющая для апреля) позволяет получить

$$\sum_{k=1}^{n} h_{k2}^2 = 0{,}25524,$$

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

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

Приступим к завершающему этапу анализа данных табл. 1 — построению интервального прогноза. Необходимо рассчитать величины $w_{ks} = c_k(t - \bar{t} - r_s) + 1/m$, если $k \in \{s + (j-1)q, j = 1, 2, ..., m\}$, 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 соотношению (18)

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

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

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

$$\Delta(13) = U(0{,}95)\sqrt{D^*[x^*(13)]} =$$

$$= 1{,}96 \cdot 385{,}7 \cdot \sqrt{0{,}61299} = 591{,}88.$$

В соответствии с оценкой (14) точечный прогноз прогностической функции:

$$x^*(13) = a^*(13 - \bar{t}) + d^* + f^*(13) =$$

$$= 212{,}26 \cdot 13 + 2587{,}48 + (-376) = 4971.$$

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

$$x_{\text{нижн}}(13) = 4971 - 592 = 4379,$$

$$x_{верх}(13) = 4971 + 592 = 5563.$$

Реальное значение — 4336 [3]. 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 $\Delta(14) = 1,96 \cdot 385,7 \cdot \sqrt{0,72480} = 643,60$. Точечный прогноз $x^*(14) = 6092$, а нижняя и верхняя доверительные границы составляют: $x_{нижн}(14) = 5448$, $x_{верх}(14) = 6736$. Реальное значение — 5430 [3]. Оно практически совпадает с нижней доверительной границей прогностической функции $x_{нижн}(14)$.

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 кажущимся невязкам (см. табл. 1). Напомним, что это распределение не является нормальным, не описывается элементом какого-либо параметрического семейства. Интервальный прогноз индивидуального значения можно 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 из нижней доверительной границы прогностической функции отнять 512, а для расчета верхней доверительной границы индивидуального значения — к верхней доверительной границе прогностической функции прибавить 326.

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

О проверке условий (7). Рассмотрим три следующих 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, $k = 1, 2, ..., n$. Более тщательно рассмотрим ранее принятую модель с целым числом периодов, для которой справедливо соотношение (11). При этом объем наблюдений $n = mq$, где m — 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 — 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 m 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 помесячные данные бухгалтерского учета. При изучении 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 выполняются. Если система точек наблюдения не образует арифметическую прогрессию, предпосылки модели также не выполняются.

В рассматриваемой модели естественно принять, что

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

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), изменив для этого (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). 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 $m \to \infty$. Таким образом, первое из условий (7) следует 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

$\{t_k, k = 1, 2, ..., n\}$ 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.

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

В соответствии с равенством (11) формулу (16) можно записать для любого $j = 1, 2, ..., n$, 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,

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

где h_{ik} — те же, что и в формуле (16). Из выражения (21) следует, что

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

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

Изменим порядок суммирования во втором слагаемом последней формулы:

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

Поскольку E_i — независимые одинаково распределенные случайные величины с математическим ожиданием нуль и дисперсией σ^2, то

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

Величину σ^2 оцениваем по формулам (10) и (15), $h_{ik(j)}$ описаны после выражения (16). Подставив оценку σ^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

$$\left| \frac{Y}{\sqrt{D^*(Y)}} \right| \le C(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \qquad (24)$$

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

Асимптотическая несмещенность оценки параметра a. Приведем пример, когда второе условие (7) не выполнено. Измерять время будем в месяцах. Возьмем данные 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 которое, 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

$$\sum_{i=1}^{n} t_i f(t_i) = 2 \cdot (-1) + 5 \cdot (-2) + 8 \cdot (-3) +$$

$$+ 11 \cdot 6 + 14 \cdot (-1) + 17 \cdot (-2) + 20 \cdot (-3) + 23 \cdot 6 =$$

$$= (-2) + (-10) + (-24) + 66 + (-14) +$$

$$+ (-34) + (-60) + 138 = 60.$$

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 составит

$$(x + 2) \cdot (-1) + (x + 5) \cdot (-2) + (x + 8)(-3) + (x + 11) \cdot 6 =$$

$$= 2 \cdot (-1) + 5 \cdot (-2) + 8 \cdot (-3) + 11 \cdot 6 = 30.$$

Причина нарушения второго условия (7) ясна — периодическая составляющая несимметрична 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 составит

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

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 место асимптотическая несмещенность оценки a^* параметра a. Если второе условие (7) выполнено, эта оценка является несмещенной, а если не выполнено — смещенной, но смещение стремится к нулю при росте числа периодов. Таким образом, выполнение второго условия (7) не является необходимым для применения рассматриваемых методов. Тем не менее проверка второго условия (7) по экспериментальным данным является полезной для решения вопроса о том, можно ли пользоваться асимптотической несмещенностью оценки при имеющемся объеме данных.

Можно заключить, что по сравнению с эвристическими алгоритмами [1] разработанная теория позволила:

1) обосновать эти алгоритмы 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 прогностической функции, трендовой и периодической составляющих, индивидуальных значений временно́го ряда.

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

В рамках математической статистики удается 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 здесь 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 работе [7]). Работа выполнена в рамках новой парадигмы прикладной статистики [8].

ЛИТЕРАТУРА

1. **Елисеева И. И., Курышева С. В., Гордеенко Н. М. и др.** Практикум по эконометрике: Учеб. пособие / Под ред. И. И. Елисеевой. — М.: Финансы и статистика, 2001. — 192 с.
2. **Орлов А. И.** Эконометрика: учебник для вузов. Изд. 4-е, доп. и перераб. — 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. — Saarbrucken: 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. Т. 78. № 1. Ч. I. С. 87 – 93.