

## Алгоритм оценки дисперсии для взвешенного метода наименьших квадратов\*

*А. В. Прасолов, Н. Г. Иванов, Н. В. Смирнов*

Санкт-Петербургский государственный университет,  
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

**Для цитирования:** *Прасолов А. В., Иванов Н. Г., Смирнов Н. В.* Алгоритм оценки дисперсии для взвешенного метода наименьших квадратов // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2023. Т. 19. Вып. 4. С. 484–496. <https://doi.org/10.21638/11701/spbu10.2023.405>

Приведено представление модели временного ряда как кусочно-стационарного процесса, т. е. набора последовательных стационарных интервалов. Для данной модели разработан алгоритм определения области, в которой расположен тренд. Известно, что определение тренда методом наименьших квадратов в чистом виде не применяется в статистическом анализе и эконометрических пакетах программ. Обычно используют взвешенный метод наименьших квадратов, чтобы в идеале убрать нестационарность. Предложен алгоритм оценки весовых коэффициентов для рассматриваемого метода с помощью кусочно-стационарного моделирования. Алгоритм апробирован на временных рядах различной природы.

*Ключевые слова:* временные ряды, кусочно-стационарный процесс, взвешенный метод наименьших квадратов.

**1. Введение.** Стохастические процессы играют важную роль в анализе и прогнозировании явлений в различных областях, таких как экономика и финансы (курсы валют, цены на нефть, стоимости акций, инфляция, валовой внутренний продукт страны), климатология (температура, атмосферное давление, скорость ветра, осадки), инженерия (вибрации машин и оборудования, акустические временные ряды, деформации, напряжения), биология (популяции, клинические параметры: уровень глюкозы в крови, кровяное давление, пульс, температура тела) и др. Важной характеристикой стохастических процессов является их стационарность, т. е. статистические свойства процесса не меняются со временем. Однако в реальных приложениях часто сталкиваются с процессами, которые не удовлетворяют условиям стационарности. Такие процессы называют нестационарными.

Важный аспект анализа нестационарных процессов — приведение их к стационарным. Это позволяет применять классические методы анализа временных рядов, разработанные для стационарных процессов. В данной работе будет рассмотрено выделение из нестационарного процесса кусочно-стационарного.

Задачи нестационарных процессов описывались Я. Р. Магнусом [1]. Он выделяет кратко-, средне- и долгосрочные горизонты прогнозирования, о которых будет говорить далее, занимается вопросами гетероскедастичности. И. С. Березин рассматривает методы аппроксимации, в том числе системы тригонометрических полиномов

\* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 20-31-90063).

© Санкт-Петербургский государственный университет, 2023

Чебышева [2]. Модели долгосрочного прогнозирования изучает Т. Андерсон [3]. Методам сглаживания посвящена работа Е. Е. Слуцкого [4].

**2. Модель временного ряда как кусочно-стационарный процесс.** Будем строить пространственно-временную область, содержащую тренд временного ряда как линию математических ожиданий. Траектория является реализацией стохастического процесса с дискретным временем. Поэтому любая ее аппроксимация случайна и не может претендовать на оценку тренда. Предлагается произвольный временной ряд трактовать как траекторию кусочно-стационарного процесса. Это позволит описать алгоритм создания области, где расположен тренд.

Рассмотрим скалярный временной ряд  $\{x_i\}_{i=\overline{1,n}}$ . Так как для каждого  $i$  число  $x_i$  есть реализация случайной величины  $X_i$  со своим распределением, то построение математической модели означает оценку тренда как линии математических ожиданий  $M(X_i)$ . В идеальном случае для прогноза необходимо знать  $M(X_i)$ ,  $i = \overline{1,n}$ . Для полноты картины хорошо бы получить и дисперсии  $D(X_i)$ . Но это невозможно, так как нам известны только реализации  $\{x_i\}$ . На практике зачастую строятся некоторая аппроксимация траектории  $\{x_i\}$  и полоса вокруг нее, внутри которой с некоторой вероятностью находится тренд. С нашей точки зрения, последнее утверждение, как правило, не доказывается [5, 6]. Самый изученный и распространенный способ моделирования основан на предположении о стационарности стохастического процесса с дискретным временем, т. е. о независимости от  $i$  функции распределения вероятностей случайной величины  $X_i$ . В этом случае применяют формулы [7]

$$\begin{aligned} X &= X_i, \\ M(X) &\approx \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \\ D(X) &\approx \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \sigma^2. \end{aligned}$$

Однако предположение о стационарности не реалистично для большинства временных рядов.

Второй широко используемый метод построения модели — разложение наблюдений на аддитивные составляющие — детерминированную и случайную:

$$x_i = u_i + \varepsilon_i,$$

где от возмущений  $\varepsilon_i$  требуется центрированность ( $M(\varepsilon_i) = 0$ ), независимость ( $M(\varepsilon_i \varepsilon_j) = 0$ ) и гомоскедастичность ( $D(\varepsilon_i) = \sigma^2$ ). Причем детерминированная часть  $u_i$  математической реализации случайной величины  $x_i$  является липшицевой с малой константой, т. е. меняется не быстро. Тогда трендом временного ряда (линией математических ожиданий) становится  $u_i$ :

$$M(X_i) = u_i.$$

Если задать класс функций  $U_i = u_i(b_1, \dots, b_m)$ , зависящих от параметров  $\{b_1, \dots, b_m\}$ , то возникает задача поиска в этом классе такого представителя (т. е. такого набора параметров), что  $\sigma^2$  будет минимальна. Решение этой задачи (единственное или нет) обязательно существует. Если параметры  $\{b_1, \dots, b_m\}$  определяют

$u_i$  линейно, то получаем известную задачу линейной регрессии. Очевидно, что оптимальные параметры  $\{b_1, \dots, b_m\}$  — это функции случайных наблюдаемых значений  $\{x_i\}$ , т. е. случайные величины, а не детерминированные, как бы хотелось. Во многих источниках [1, 8–10] класс функций задается на основе некоторого знания природы стохастического процесса. Обычно выделяют монотонную составляющую (и называют ее трендом), несколько видов сезонных колебаний, частоты которых специально подбираются, и т. д. Примерно по такой же схеме работает метод Холта — Уинтерса [11].

Проведем объединение (в некотором смысле) двух описанных выше подходов. Моделирование временного ряда осуществляется в два этапа. На первом этапе рассматривается предположение о стационарности стохастического процесса на некотором (коротком, но длиннее 5 шагов) временном интервале. Вычисляются выборочное среднее и выборочное среднеквадратическое отклонение. Оценивается дисперсия статистик. Добавляется следующее наблюдение (следующий элемент временного ряда) и проверяется гипотеза о принадлежности его генеральной совокупности, определенной наблюдениями в предшествующие моменты времени (точнее говоря, генеральной совокупности, выборкой которой явились предшествующие временные узлы). Если очередная реализация не отвергает гипотезу, то объединяем выборку и последнее наблюдение. Затем повторяем процедуру, пока новое наблюдение не отвергает гипотезу о принадлежности одной генеральной совокупности. Факт отказа от гипотезы закрывает процедуру удлинения интервала стационарности. Для него вычисляем выборочное среднее и оценку разброса при заданной гарантии надежности.

Далее снова повторим всю процедуру, пока не будут исчерпаны все наблюдения с  $i = \overline{1, n}$ . В результате получим кусочно-постоянную функцию и зону вокруг нее, где обязательно находится тренд всего временного ряда. Таким образом, осуществляется переход к кусочно-стационарному стохастическому процессу.

**Замечание 1.** Решение выбрать базовый интервал стационарности именно из 5 значений было принято из следующих соображений: одно значение рассматривать невозможно, так как не определить дисперсию  $D(x)$ , ведь в знаменателе  $n - 1$ , два значения могут с большой вероятностью находиться по одну сторону от математического ожидания, три и более значений рассматривать уже адекватно. Теория малых выборок предполагает выбор от 5 до 30 значений. Таким образом, было принято решение за базовый кусочно-стационарный интервал взять минимально возможное значение — 5 наблюдений.

**2.1. Алгоритм построения интервала стационарности.** Обсудим детально алгоритм построения интервала стационарности. Предположим, что предыдущий интервал закончился номером  $i$  ( $i \in [1, n - 5]$ ). Рассмотрим наблюдения  $x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}, x_{i+5}$ . Относительно этих чисел будем считать, что они суть реализации случайной величины  $X$  с математическим ожиданием  $a$  и дисперсией  $\sigma^2$ , т. е. на номерах  $k = \overline{i + 1, i + 5}$  временной ряд ведет себя как стационарный. Величины  $a$  и  $\sigma^2$  неизвестны, но, предполагая распределение  $X$  нормальным, получим несмещенные и состоятельные оценки величин  $a$  и  $\sigma^2$  соответственно [7]:

$$\bar{x} = \frac{1}{5} \sum_{k=i+1}^{i+5} x_k, \quad \hat{s}^2 = \frac{1}{5} \sum_{k=i+1}^{i+5} (x_k - \bar{x})^2.$$

Отметим, что при малых объемах выборки  $\bar{x}$  не обязательно распределено нормально,

поэтому используем распределение Стьюдента. Введем статистику

$$t = \frac{\bar{x} - a}{s} \sqrt{n}.$$

По таблицам, приведенным во многих источниках, определим критическое значение  $t_{\gamma, n-1}$ , при котором

$$P\left(\left|\frac{\bar{x} - a}{s} \sqrt{n}\right| < t_{\gamma, n-1}\right) = \theta(t, n-1) = \gamma.$$

Заметим, что если ограничиться надежностью  $\gamma = 0.95$  (95 %), то можно приближенно оценить  $t_{\gamma, k} \approx 2 + \frac{2.63}{k}$ ,  $k = 4, 5, 6, \dots$ , и тем самым избежать многочисленных обращений к таблицам. Для надежности  $\gamma = 0.9$  (90 %) приближенная формула имеет вид  $t_{\gamma, k} \approx 1.64 + \frac{1.82}{k}$ ,  $k = 4, 5, 6, \dots$ . И соответственно для  $\gamma = 0.99$  (99 %) приближенная формула будет  $t_{\gamma, k} \approx 2.4 + \frac{8.15}{k}$ ,  $k = 4, 5, 6, \dots$ .

Доверительный интервал для генеральной средней  $a$  находится по формуле

$$\bar{x} - \Delta \leq a \leq \bar{x} + \Delta,$$

где  $\Delta = \frac{t_{\gamma, n-1} s}{\sqrt{n}}$ , или в случае  $\gamma = 0.95$  используем приближенную формулу

$$\Delta \approx \left(2 + \frac{2.63}{n-1}\right) \frac{\hat{s}}{\sqrt{n}}, \quad n = 5, 6, 7, \dots$$

Далее рассмотрим очередное наблюдение временного ряда и проверим, сохраняется ли предположение о стационарности стохастического процесса, т. е. будет ли очередное наблюдение принадлежать выборке генеральной совокупности, которая определилась предшествующими наблюдениями. Таким образом, включая все новые наблюдения, формируем интервал стационарности временного ряда. Ответ на вопрос, поставленный выше, дается известными алгоритмами оценивания, принадлежат ли две выборки одной генеральной совокупности с заданной надежностью  $\gamma$ . В данной работе увеличение интервала стационарности останавливается, как только появится наблюдение, которое «выпадает» из получившейся до этого выборки, т. е. оно из другой генеральной совокупности. В современной математической статистике есть много алгоритмов, как определить «выпадение» [7]. Один из них рассматривает новое значение как вторую выборку в задаче определения однородности двух выборок с заданной надежностью. Опишем его.

Пусть имеется выборка  $x_1, x_2, \dots, x_n$  из нормально распределенной генеральной совокупности с неизвестными  $a$  и  $\sigma^2$ . Оценим их с помощью  $\bar{x}$  и  $\hat{s}^2$ . Добавим из временного ряда следующее наблюдение  $x^*$ . И будем рассматривать его как выборку с объемом 1. Если опираться на предположение о нормальности всех изучаемых распределений, то и при малых объемах выборок статистика  $t = \frac{\bar{x} - x^*}{s} \sqrt{n}$  имеет  $t$ -распределение Стьюдента с  $k = n - 1$  степенями свободы. И, следовательно, гипотеза о принадлежности  $x^*$  той же самой генеральной совокупности отвергается, если  $|t| \geq t_{\gamma, n-1}$ . Считаем, что в этом случае предположение о стационарности отвергается. Выбираем следующий набор наблюдений (например, из 5 следующих членов временного ряда) и проделываем ту же процедуру. В результате временной ряд создает кусочно-постоянную функцию, соответствующую «кусочной» стационарности стохастического процесса. Полоса, определяемая неравенством  $\bar{x} - \Delta \leq a \leq \bar{x} + \Delta$ , вокруг этой функции обязательно содержит тренд стохастического процесса. И, кроме

того, сам временной ряд почти всегда лежит внутри этой полосы с надежностью  $\gamma$ . Интервалы стационарности и скачки (разности соседних значений средних) составляют случайную последовательность.

**2.2. Пример.** Рассмотрим реальные недельные данные цены нефти марки Brent за 2015 г. (рис. 1) [12].

Месяц	Неделя 1		Неделя 2		Неделя 3		Неделя 4		Неделя 5	
Январь	02.01	56.03	09.01	49.47	16.01	46.58	23.01	46.44	30.01	46.76
Февраль	06.02	54.62	13.02	56.57	20.02	60.57	27.02	60.63		
Март	06.03	60.12	13.03	56.51	20.03	52.72	27.03	55.01		
Апрель	03.04	54.79	10.04	56.51	17.04	59.12	24.04	61.41		
Май	01.05	63.49	08.05	65.01	15.05	64.90	22.05	64.31	29.05	61.57
Июнь	05.06	61.90	12.06	63.24	19.06	60.65	26.06	60.84		
Июль	03.07	60.36	10.07	56.63	17.07	57.17	24.07	55.75	31.07	54.14
Август	07.08	48.59	14.08	47.94	21.08	46.00	28.08	43.53		
Сентябрь	04.09	48.87	11.09	47.60	18.09	47.34	25.09	47.33		
Октябрь	02.10	46.99	09.10	51.33	16.10	48.99	23.10	46.81	30.10	47.15
Ноябрь	06.11	47.23	13.11	44.75	20.11	41.54	27.11	43.65		
Декабрь	04.12	42.41	11.12	38.76	18.12	36.84	25.12	35.90		

Рис. 1. Таблица недельных данных марки Brent в 2015 г. (в долл. США)

Известно, что это плохие для прогнозирования данные: нет периодических или монотонных составляющих, ярко выражена гетероскедастичность и пр.

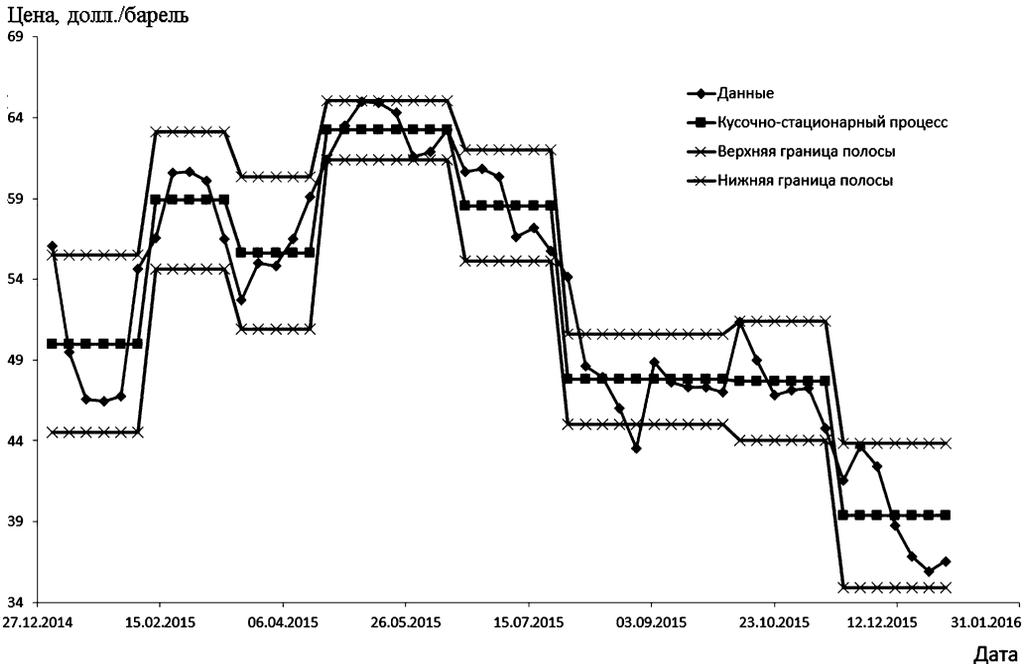


Рис. 2. Цена нефти марки Brent

На рис. 2 видны интервалы стационарности, полученные описанным алгоритмом. На каждом интервале извилистая линия изображает реальные цены нефти, горизон-

тальная соответствует выборочному среднему, от которого вверх и вниз откладываются пределы доверительного интервала (в данном случае с надежностью 99 %). Доверительная полоса, образованная доверительными интервалами по всей длине временного ряда, обязательно содержит тренд (линию математических ожиданий) стохастического процесса. Отметим, что если бы доверительная полоса строилась вокруг временного ряда как траектории стохастического процесса, то она не гарантировала бы содержание тренда.

**3. Оценка дисперсии взвешенного метода наименьших квадратов.** Продолжим исследования, связанные с полученным алгоритмом выделения кусочно-стационарного процесса из нестационарного.

При анализе временных рядов в реальной жизни часто сталкиваются с тем, что ряды являются нестационарными процессами, обладающими гетероскедастичностью. Напомним, что если дисперсия случайной ошибки не является постоянной:  $D(\varepsilon_i) = \sigma_i^2 \neq \text{const}$  и ошибки некоррелированы, то такая ситуация называется гетероскедастичностью.

Для получения эффективных (оптимальных) оценок параметров можно воспользоваться так называемым взвешенным методом наименьших квадратов (МНК). Чтобы понять, как он работает, рассмотрим случай, когда дисперсия случайных ошибок  $D(\varepsilon_i) = \sigma_i^2$  известна [13].

Пусть рассматривается модель

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \dots + \beta_k x_i^{(k)} + \varepsilon_i,$$

для которой выполнены все предпосылки классической линейной модели множественной регрессии за одним исключением: в данных наблюдается гетероскедастичность  $D(\varepsilon_i) = \sigma_i^2$ . В таком случае можно разделить правую и левую части уравнения регрессии на  $\sigma_i$ :

$$\frac{y_i}{\sigma_i} = \frac{\beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \dots + \beta_k x_i^{(k)} + \varepsilon_i}{\sigma_i}.$$

После этого сделаем простую замену переменных:

$$\tilde{y}_i = \frac{y_i}{\sigma_i}, \quad \tilde{x}_i^{(1)} = \frac{1}{\sigma_i}, \quad \tilde{x}_i^{(2)} = \frac{x_i^{(2)}}{\sigma_i}, \dots, \quad \tilde{x}_i^{(k)} = \frac{x_i^{(k)}}{\sigma_i}, \quad \tilde{\varepsilon}_i = \frac{\varepsilon_i}{\sigma_i}.$$

В результате переходим к новой модели:

$$\tilde{y}_i = \beta_1 \tilde{x}_i^{(1)} + \beta_2 \tilde{x}_i^{(2)} + \dots + \beta_k \tilde{x}_i^{(k)} + \tilde{\varepsilon}_i.$$

Она полезна тем, что в ней гетероскедастичности нет, так как дисперсия случайной ошибки является константой:

$$D(\tilde{\varepsilon}_i) = D\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2} D(\varepsilon_i) = \frac{1}{\sigma_i^2} \sigma_i^2 = 1 = \text{const}.$$

Следовательно, МНК, примененный к новой модели, будет давать эффективный результат. Таким образом, суть взвешенного МНК состоит в том, чтобы сделать правильную замену переменных так, чтобы применение к новой модели (с измененными переменными) обычного МНК приводило к получению эффективных оценок коэффициентов. После этого для интерпретации результатов можно вернуться к исходным переменным.

Поясним, почему МНК называется взвешенным. В случае взвешенного МНК минимизируем с помощью  $\beta_i$  сумму квадратов остатков  $e_i^2$  новой модели:

$$\sum_{i=1}^n \frac{1}{\sigma_i^2} e_i^2 \rightarrow \min_{\hat{\beta}}$$

Получается, что минимизируем сумму квадратов остатков, но каждое слагаемое домножается на весовой коэффициент  $\frac{1}{\sigma_i^2}$ , т. е. минимизируем сумму квадратов остатков с определенными весами. Чем меньше дисперсия для  $i$ -го наблюдения (т. е. чем меньше для него фактор случайности), тем больший вес это наблюдение имеет в той сумме, которую минимизируем. Тем самым наибольший вес придаем наиболее «надежным» наблюдениям, что и позволяет улучшить качество получаемых оценок.

Разумеется, в реальности дисперсия случайной ошибки обычно неизвестна исследователю, что приводит к необходимости рассмотрения более реалистичного случая, когда надо сначала оценить  $\sigma_i^2$ , а затем эту оценку применять к взвешенному МНК.

**Замечание 2.** Взвешенный МНК предпочтительней стандартного, так как он дает возможность получить эффективные оценки коэффициентов, но только при условии, что правильно специфицировано уравнение для дисперсии случайной ошибки (т. е. правильно поняли, как именно устроена гетероскедастичность в анализируемой модели).

В частном случае, когда дисперсия случайной ошибки прямо пропорциональна квадрату единственной переменной  $b$ :  $D(\varepsilon_i) = \sigma_i^2 = \sigma_i^2 b_i^2 > 0$ , приводится алгоритм реализации взвешенного МНК [1, 13]. Если дисперсия случайной ошибки неизвестна, то некоторыми исследователями предлагается оценить сначала исходную модель с помощью МНК и получить остатки  $e_i$ , а затем вспомогательную модель для остатков, где регрессорами вспомогательной модели являются регрессоры исходной модели, а также их квадраты (переменные, которые влияют на дисперсию случайной ошибки) [13]. Однако допущение о том, что модель остатков на всех данных описывается единой функцией, не очень хорошо согласуется с нестационарностью процессов в реальных данных, когда выбросы могут быть разными для различных этапов процесса.

**3.1. Алгоритм.** Предложим разработанный алгоритм взвешенного МНК с оценкой  $\sigma_i^2$  для исходного временного ряда  $X(T_i), i = \overline{1, n}$ :

1) согласно природе данных, специфицируем класс функций  $Y(T_i)$  для аппроксимации временного ряда;

2) применим к ряду  $X(T_i)$  МНК, получая в результате  $y(T_i)$  — представителя класса  $Y(T_i)$ ;

3) вычислим остатки  $e(T_i) = X(T_i) - y(T_i)$ ;

4) для ряда остатков  $e(T_i)$  выделим кусочно-стационарный процесс с параметрами  $M(T_i)$  — временной кусочно-стационарный ряд,  $\Delta(T_i)$  — кусочно-постоянное отклонение для  $M(T_i)$ , которое и будем считать оценкой  $\sigma_i$ ;

5) применим для ряда  $y(T_i)$  взвешенный МНК с  $\sigma_i = \Delta(T_i), i = \overline{1, n}$ ;

6) получим новую модель  $\tilde{y}(T_i)$ . Делаем обратную замену переменных:  $z(T_i) = \tilde{y}(T_i)\sigma_i, i = \overline{1, n}$ . В итоге имеем  $z(T_i)$  — модель аппроксимации ряда  $X(T_i)$  с помощью взвешенного МНК.

**3.2. Примеры.** Во всех примерах будем использовать уровень надежности для кусочно-стационарного процесса,  $\gamma = 0.95$ . В основном в них будет проводиться спецификация линейной составляющей и сезонной в виде пары синуса и косинуса, причем

период у сезонных составляющих не равен длине всего интервала идентификации. В данном случае прогноз будет продолжением функции.

*Пример 1.* Рассмотрим ряд данных  $X(T_i) = 0.25 \sin(T_i) + 0.75 \cos(T_i)$ ,  $i = \overline{1, 100}$ ,  $T_i = \frac{i}{5}$ . Для аппроксимации будем искать представителя из класса функций  $Y(T_i) = \beta_1 + \beta_2 T_i + \beta_3 \cos(T_i) + \beta_4 \sin(T_i)$ . Применяя алгоритм взвешенного МНК, получим для него меньшую сумму квадратов отклонений ( $1.106 \cdot 10^{-30}$ ), чем для стандартного МНК ( $1.2087 \cdot 10^{-30}$ ) (рис. 3), причем накопительная сумма разностей  $e(T_i) - (X(T_i) - z(T_i))$  растет без сильных скачков на протяжении всех  $T_i = \frac{i}{5}$  (рис. 4).

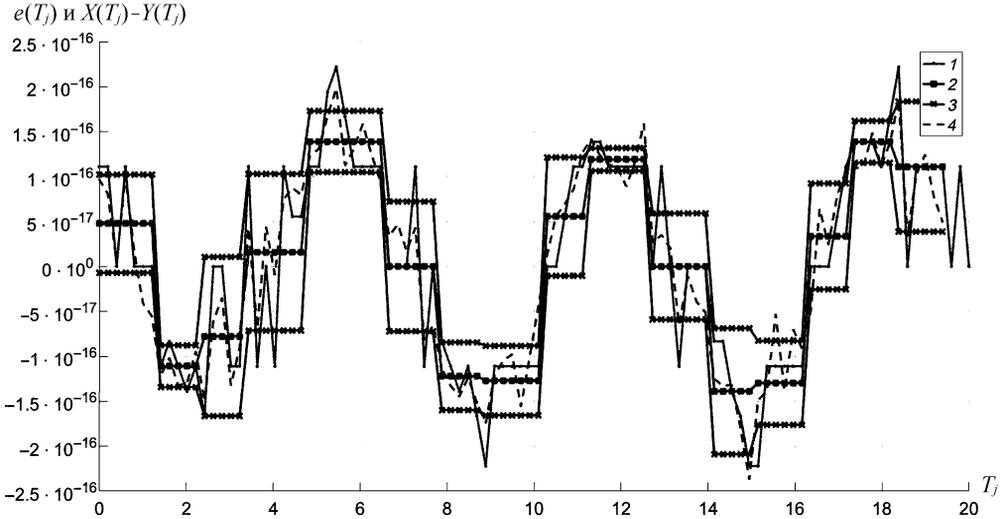


Рис. 3. Остатки двух моделей

1 —  $e(T_i)$ ; 2 — кусочно-стационарный процесс для  $e(T_i)$ ; 3 — верхняя и нижняя границы доверительного интервала; 4 —  $X(T_i) - z(T_i)$ .

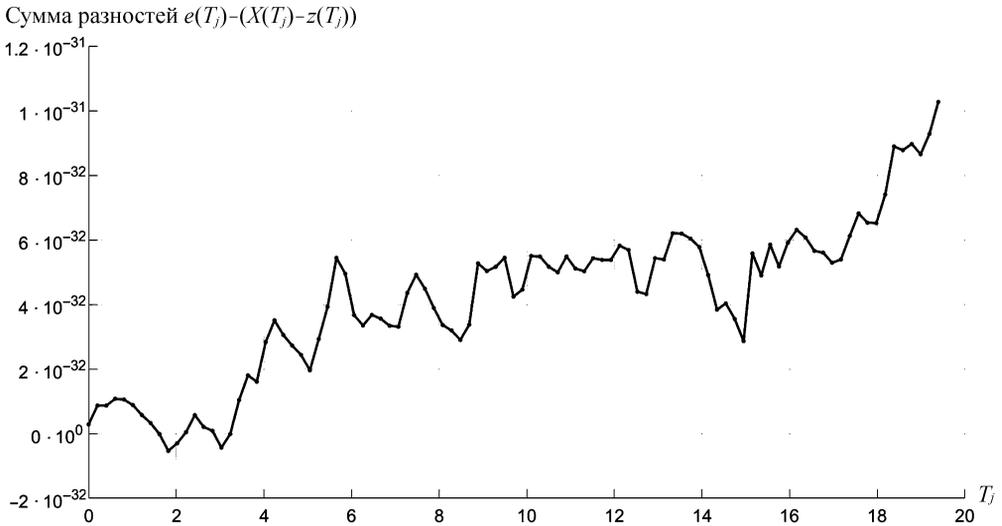


Рис. 4. Накопительная сумма разностей остатков двух моделей

*Пример 2.* Теперь рассмотрим прямую  $X(T_i) = 2 + 0.3t$ ,  $i = \overline{1, 100}$ ,  $T_i = \frac{i}{5}$ . Для аппроксимации будем искать представителя из того же класса функций  $Y(T_i) = \beta_1 + \beta_2 T_i + \beta_3 \cos(T_i) + \beta_4 \sin(T_i)$ . На рис. 5 показано, что даже если МНК в связи с особенностями численных вычислений ошибочно определил  $\beta_3$  и  $\beta_4$  ненулевыми, то взвешенный МНК сделает их более близкими к нулю, т. е. значительно снизит сумму квадратов остатков относительно стандартного МНК.

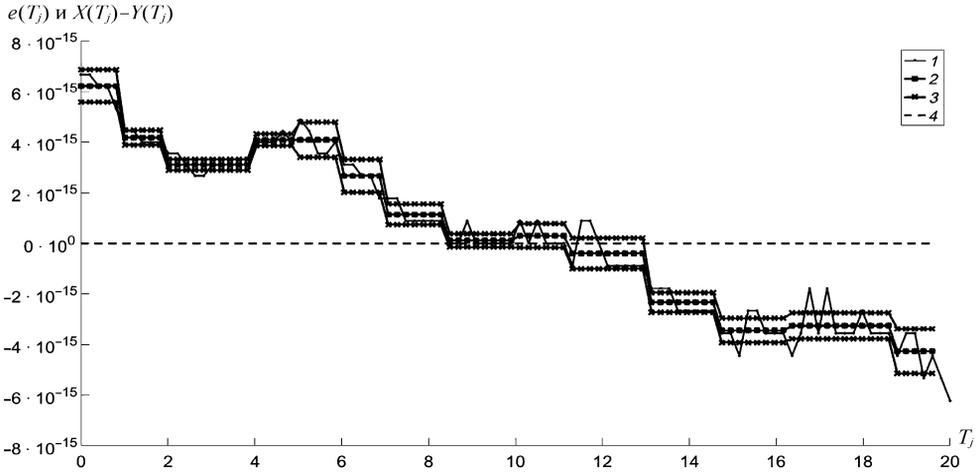


Рис. 5. Остатки двух моделей

1 —  $e(T_i)$ ; 2 — кусочно-стационарный процесс для  $e(T_i)$ ; 3 — верхняя и нижняя границы доверительного интервала; 4 —  $X(T_i) - z(T_i)$ .

*Пример 3.* Покажем, что будет, если неудачно проведена спецификация модели. Для ряда  $X(T_i) = 5 \sin(0.3t^2)$ ,  $i = \overline{1, 100}$ ,  $T_i = \frac{i}{5}$ , будем проводить аппроксимацию из класса функций  $Y(T_i) = \beta_1 + \beta_2 T_i + \beta_3 \cos(T_i) + \beta_4 \sin(T_i)$ . На рис. 6 видно, что даже близкой аппроксимации не получится из-за некорректной спецификации. Однако из

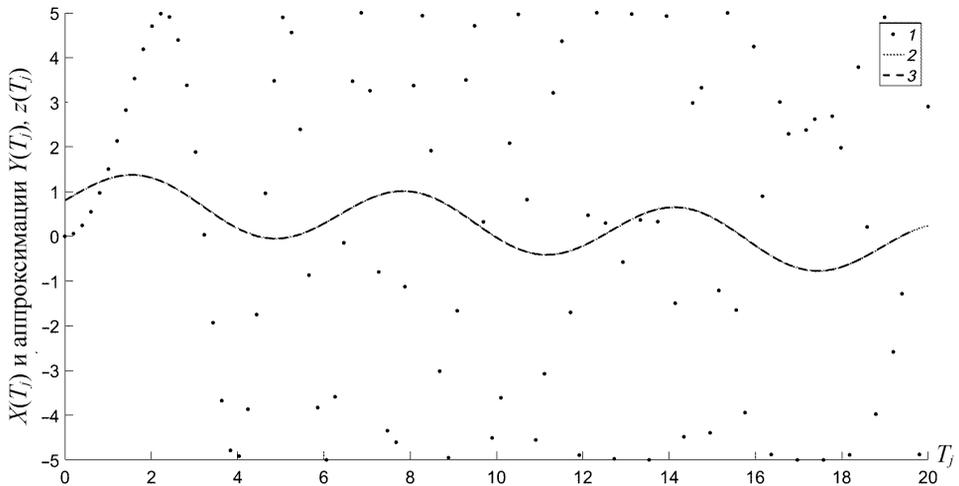


Рис. 6. Ряд  $5 \sin(0.3t^2)$  и аппроксимации

1 — реальные данные  $X(T_i)$ ; 2 —  $Y(T_i)$ ; 3 —  $z(T_i)$ .

рис. 7 следует, что для первых  $T_i$  аппроксимация достаточно близка к исходному ряду, затем и для МНК, и для взвешенного МНК отклонения становятся больше. В итоге взвешенный МНК имеет большую сумму квадратов остатков, чем стандартный МНК.

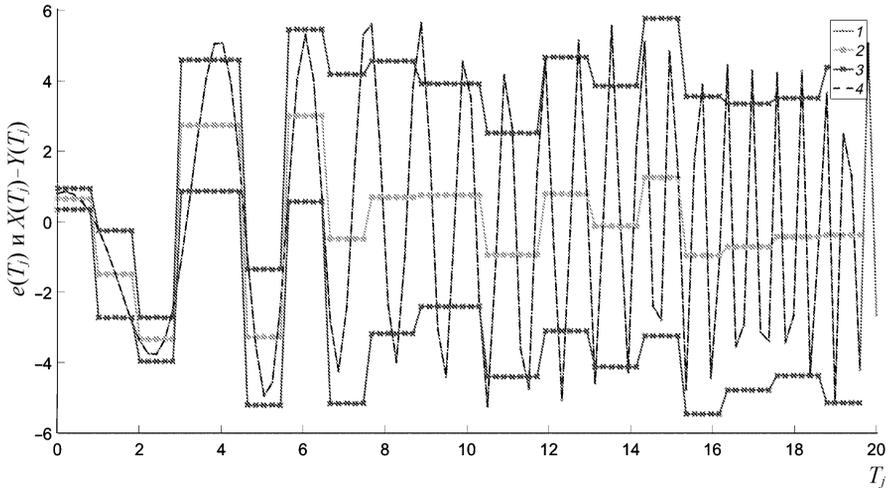


Рис. 7. Остатки двух моделей

1 —  $e(T_i)$ ; 2 — кусочно-стационарный процесс для  $e(T_i)$ ; 3 — верхняя и нижняя границы доверительного интервала; 4 —  $X(T_i) - z(T_i)$ .

*Пример 4.* Наконец, приведем пример реальных данных. Это данные о ежедневных замерах температуры в Москве [14]. Аппроксимировать временной ряд  $X(T_i)$ ,  $i = \overline{1,500}$ ,  $T_i = i$ , будем немного другим классом функций  $Y(T_i) = \beta_1 + \beta_2 T_i + \beta_3 \cos(\beta_4 T_i) + \beta_5 \sin(\beta_6 T_i)$ . Его достаточно, чтобы аппроксимировать годовую сезонность. Для чего введем амплитуды  $\beta_4$  и  $\beta_6$ . Однако стандартным МНК можно определить только  $\beta_1, \beta_2, \beta_3, \beta_5$ . Чтобы определить  $\beta_1, \dots, \beta_6$ , применим численный нелинейный метод Левенберга — Марквардта [15], решающий задачу наименьших квадратов. Принцип метода является методом спуска: задаются начальные коэффициенты, и алгоритм приближает к исходному ряду, пока не будет выполнено условие останова либо достигнуто максимально заданное число итераций. В результате (рис. 8) видно, что спецификация модели недостаточно точная, остатки  $e(T_i)$  достаточно большие, поэтому взвешенный МНК дал практически тот же результат, что и стандартный МНК.

**4. Заключение.** В работе приведен алгоритм выделения из нестационарного временного ряда кусочно-стационарного. Для него построена полоса, внутри которой с определенной вероятностью лежит любая из реализаций данного процесса. С помощью предложенного алгоритма оценена дисперсия для взвешенного МНК, который дает эффективный результат, убирая гетероскедастичность. Однако в реальной жизни  $\sigma$  необходимо оценить. Этот метод показывает, что для некоторых рядов метод действительно помогает сделать более точную аппроксимацию при должной спецификации модели. В то же время при плохой спецификации модели и из-за особенностей алгоритма построения кусочно-стационарного процесса взвешенный МНК может дать результат даже хуже стандартного. Поэтому необходимо понимать природу ряда и разумно применять рассмотренный метод.

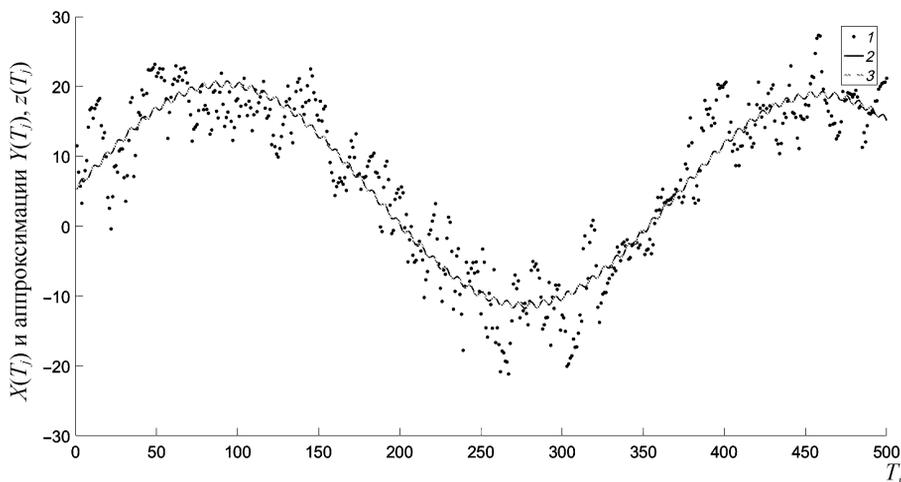


Рис. 8. Температура в Москве и аппроксимации  
 1 — реальные данные  $X(T_i)$ ; 2 —  $Y(T_i)$ ; 3 —  $z(T_i)$ .

Все вычисления выполнены с помощью MS Excel<sup>®</sup> и прикладного вычислительного пакета MATLAB<sup>®</sup>.

## Литература

1. Магнус Я. Р., Катышев П. К., Персецкий А. А. Эконометрика. М.: Дело, 2004. 575 с.
2. Березин И. С., Жидков Н. П. Методы вычислений. В 2-х т. Т. 1. Изд. 2-е, стереотип. М.: Физматлит, 1962. 464 с.
3. Андерсон Т. Статистический анализ временных рядов / пер. с англ. И. Г. Журбенко, В. П. Носко; под ред. Ю. К. Беляева. М.: Мир, 1976. 757 с.
4. Слуцкий Е. Е. Избранные труды. Теория вероятностей. Математическая статистика. М.: Изд-во АН СССР, 1960. 292 с.
5. Иванов Н. Г., Прасолов А. В. Анализ аппроксимации тренда временного ряда и прогнозирования на его основе // Устойчивость и процессы управления. Памяти В. И. Зубова (SCP). СПб.: Издательский Дом Федоровой Г. В., 2015. С. 423–424.
6. Прасолов А. В. Некоторые математические методы экономической теории. СПб.: НОВА, 2016. 286 с.
7. Кремер Н. Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов, обучающихся по экономическим специальностям М.: ЮНИТИ-ДАНА, 2012. 551 с.
8. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление / пер. с англ. А. Л. Левшина; под ред. В. Ф. Писаренко. М.: Мир, 1974. 406 с.
9. Поджорытова О. А., Соколов М. В. Анализ временных рядов: учеб. пособие для бакалавриата и магистратуры. М.: Юрайт, 2016. 267 с.
10. Буре В. М., Париллина Е. М. Теория вероятностей и математическая статистика. СПб.: Лань, 2013. 416 с.
11. Домбровский В. В. Методы количественного анализа финансовых операций. Томск: Изд-во научно-технической литературы, 1998. 101 с.
12. Economic Research. Federal Reserve bank of St. Louis. URL: <https://research.stlouisfed.org/fred2/series/POILBREUSDM> (дата обращения: 3 августа 2023 г.).
13. Картаев Ф. С. Дружелюбная эконометрика. URL: <https://books.econ.msu.ru/Introduction-to-Econometrics/> (дата обращения: 3 августа 2023 г.).
14. Average Daily Temperature Archive. URL: <http://academic.udayton.edu/kissock/http/Weather/gsd95-current/RSMOSCOW.txt> (дата обращения: 3 августа 2023 г.).
15. Function Documentation lsqonlin. URL: <https://www.mathworks.com/help/optim/ug/lsqonlin.html> (дата обращения: 3 августа 2023 г.).

Статья поступила в редакцию 25 августа 2023 г.  
Статья принята к печати 12 октября 2023 г.

Контактная информация:

Прасолов Александр Витальевич — д-р физ.-мат. наук, проф.; a.prasolov@spbu.ru

Иванов Никита Григорьевич — st013355@student.spbu.ru

Смирнов Николай Васильевич — д-р физ.-мат. наук, проф.; n.v.smirnov@spbu.ru

## Algorithm of variance estimation in weighted least squares method\*

A. V. Prasolov, N. G. Ivanov, N. V. Smirnov

St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg,  
199034, Russian Federation

**For citation:** Prasolov A. V., Ivanov N. G., Smirnov N. V. Algorithm of variance estimation in weighted least squares method. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2023, vol. 19, iss. 4, pp. 484–496.  
<https://doi.org/10.21638/11701/spbu10.2023.405> (In Russian)

The representation of a time series model as a piecewise-stationary process is provided, wherein it is regarded as a collection of successive stationary intervals. An algorithm has been developed for identifying the domain containing the trend within this model. It is recognized that applying the least squares method directly for trend determination is not commonly employed in statistical analysis and econometric software packages. Typically, the weighted least squares method is utilized to ideally eliminate non-stationarity. The authors presents an algorithm for estimating the weight coefficients for this method through piecewise-stationary modeling. The algorithm has been tested on time series of various natures.

*Keywords:* time series, piecewise-stationary process, weighted least squares method.

## References

1. Magnus Ya. R., Katyshev P. K., Persetsky A. A. *Ekonometrika [Econometrics]*. Moscow, Delo Press, 2004, 575 p. (In Russian)
2. Berezin I. S., Zhidkov N. P. *Metody vychislenij [Computational methods]*. In 2 vols, vol. 1. Ed. 2<sup>nd</sup>, stereotype. Moscow, Fizmatlit Publ., 1962, 464 p. (In Russian)
3. Anderson T. *Statisticheskij analiz vremennyh ryadov [Statistical analysis of time series]*. Moscow, Mir Publ., 1976, 757 p. (In Russian)
4. Slutsky E. E. *Izbrannye trudy. Teoriya veroyatnostej. Matematicheskaya statistika [Selected works. Probability theory. Mathematical statistics]*. Moscow, USSR Academy of Sciences Press, 1960, 292 p. (In Russian)
5. Ivanov N. G., Prasolov A. V. Analiz approksimacii trenda vremennogo ryada i prognozirovaniya na ee osnove [Analysis of the approximation of time series trend and forecasting based on it]. *Stability and Control Processes. Memory of V. I. Zubov (SCP)*. St. Petersburg, Izdatel'skii dom Fedorovoi G. V. Publ., 2015, pp. 423–424. (In Russian)
6. Prasolov A. V. *Nekotorye matematicheskie metody ekonomicheskoi teorii [Some mathematical methods in economic theory]*. St. Petersburg, NOVA Publ., 2016, 286 p. (In Russian)
7. Kremer N. Sh. *Teoriya veroyatnostej i matematicheskaya statistika. Uchebnik dlya studentov vuzov, obuchayushchihhsya po ekonomicheskim special'nostyam [Probability theory and mathematical statistics. A textbook for university students in economics specializations]*. Moscow, UNITY-DANA Publ., 2012, 551 p. (In Russian)

---

\* This work was supported by the Russian Foundation for Basic Research (project N 20-31-90063).

8. Boks J., Jenkins G. *Analiz vremennyh ryadov. Prognoz i upravlenie* [Time series analysis. Forecasting and control]. Moscow, Mir Publ., 1974, 406 p. (In Russian)

9. Podkorytova O. A., Sokolov M. V. *Analiz vremennyh ryadov. Uchebnoe posobie dlya bakalavriata i magistratury* [Time series analysis. A study guide for bachelor's and master's programs]. Moscow, Yurayt Publ., 2016, 267 p. (In Russian)

10. Bure V. M., Parilina E. M. *Teoriya veroyatnostej i matematicheskaya statistika* [Probability theory and mathematical statistics]. St. Petersburg, Lan' Publ., 2013, 416 p. (In Russian)

11. Dombrovsky V. V. *Metody kolichestvennogo analiza finansovyh operacij* [Quantitative analysis methods for financial operations]. Tomsk, Publisher Scientific and Technical Literature, 1998, 101 p. (In Russian)

12. *Economic Research. Federal Reserve bank of St. Louis*. Available at:

<https://research.stlouisfed.org/fred2/series/POILBREUSDM> (accessed: August 3, 2023).

13. Kartaev F. S. *Druzhelyubnaya ekonometrika* [Friendly econometrics]. Available at:

<https://books.econ.msu.ru/Introduction-to-Econometrics/> (accessed: August 3, 2023). (In Russian)

14. *Average Daily Temperature Archive*. Available at:

<http://academic.udayton.edu/kissock/http/Weather/gso95-current/RSMOSCOW.txt> (accessed: August 3, 2023). (In Russian)

15. *Function Documentation lsqnonlin*. Available at:

<https://www.mathworks.com/help/optim/ug/lsqnonlin.html> (accessed: August 3, 2023)

Received: August 25, 2023.

Accepted: October 12, 2023.

#### Authors' information:

*Aleksander V. Prasolov* — Dr. Sci. in Physics and Mathematics, Professor; [a.prasolov@spbu.ru](mailto:a.prasolov@spbu.ru)

*Nikita G. Ivanov* — [st013355@student.spbu.ru](mailto:st013355@student.spbu.ru)

*Nikolay V. Smirnov* — Dr. Sci. in Physics and Mathematics, Professor; [n.v.smirnov@spbu.ru](mailto:n.v.smirnov@spbu.ru)